University of Alexandria
Faculty of Arts
Department of phonetics
and linguistics
2015

Speaker Identification based on Temporal Parameters

التعرف على المتكلم اعتماداً على معايير السرعة الزمنية

A thesis submitted to the Phonetics and Linguistics Department of the University of Alexandria in partial fulfillment of the requirements for the Degree of Master of Arts in phonetic sciences.

By/

Eman Muhammad Yousri Ragab Abdel Azim El Gamal

Under the auspices of /

Prof. Dr. Mervat Fashal
Professor of phonetic sciences
Department of phonetics and linguistics
Faculty of Arts
University of Alexandria

مكتبة وملتقى علم الأصوات
phonetics-acoustics.blogspot.com

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

https://phonetics-acoustics.blogspot.com

قال الله تعالى:

"وَقُلِ اعْمَلُوا فَسَيَرَى اللَّهُ عَمَلَكُمْ وَرَسُولُهُ وَالْمُؤْمِنُونَ وَسَتُرَدُّونَ إِلَى عَالِمِ الْغَيْبِ وَالشَّهَادَةِ فَيُنَبِّئُكُمْ بِمَا كُنتُمْ تَعْمَلُونَ."

سورة التوبة : آيه 105

https://phonetics-acoustics.blogspot.com

Dedicated to

The soul of my dad

May his soul rest in peace

https://phonetics-acoustics.blogspot.com

https://phonetics-acoustics.blogspot.com

Acknowledgments

My passion for human voices started eight years ago, since I was attending the oriental music concerts at Alexandria Opera House; the differences between singers' voices while singing and while talking were like my first main motivation. Then my passion moved on to the surrounding individuals' voices and their differences through every day conversations and phone talks with familiar and unfamiliar speakers. Friends and family members even unfamiliar listeners (who listen to my voice through phone for the first time) can recognize my voice through phone calls from the first two or three spoken syllables; moreover, they can correctly estimate my instant emotion state, my physical state and the place of the call as well; that makes me wonder how they even know all of these just from saying "hello". Last but not least, my passion for the study of human voice moved finally to academic study, specially the use of human voices for forensic investigations, which was the subject of my graduation project, and which has become a very excited discipline for me.

First and foremost, I would like to express my deepest gratitude and sincere appreciation to Prof. Dr. Mervat Fashal, Professor of Phonetic Sciences, Department of Phonetics and Linguistics, Faculty of Arts, University of Alexandria, for guiding me throughout my work on this discipline (forensic speaker identification) and for giving me the honor of supervising the present thesis. I'm greatly indebted for her unflinching effort through her strict observations, valuable guidance and suggestions that improved the outlook and organization of the thesis as it must be.

Also, I would like to make a special thanks to Prof. Dr. Hanaa Abdel Fatah, Department of Phonetics and Linguistics, Faculty of Arts, University of Alexandria, and Prof. Dr. Noha Qurany Othman, Department of Communication, Faculty of Engineering, University of Alexandria, for their kind approval to be the examiners of my work.

Second, I want to make a special mention of Prof. Dr. Muhammad Hussein, Professor of Statistics, Department of Mathematics, Faculty of Science, University of Alexandria, for his kind and generous help in presenting all the results, as it should be, through using Microsoft office excel. He was extremely magnanimous, patient with me and really taught me a lot.

Third, I would like to mention in particular Ms. Ola Nouh, who is in charge of the library of Phonetics and Linguistics department, for her serious and helpful support in facilitating all the references I need. Also I should not forget to express my deepest thanks and gratitude to everyone at the department of Phonetics and Linguistics for inspiring and encouraging me in my difficult times.

Fourth, I would like to make a special reference to every person at Alexandria Radio (Bakous, Alexandria) for offering me their kind and honorable aid especially during collecting and recording my data.

After all, forever and ever; I want to give a special thanks to my entire family for their constant support and for raising my spirit all the time. The care I received from every member of my dear family, specially my beloved mother, is simply beyond any possible attempt of description. Their loving care and their patience always made me realize how much I am blessed. Also I shouldn't forget to thank Nirmeen El-Dala' for her effort in reading and reviewing the manuscript of my thesis and for her insightful comments.

And customarily, of course, any mistakes, flaws or signs of incoherence are mine and mine alone.

List of Abbreviations

Abbreviation	Meaning
AR:	Articulation Rate
F_0:	Fundamental Frequency
FSI:	Forensic Speaker Identification
S.S.: **syll./ sec.:**	Syllable / Second
SR:	Speech Rate

https://phonetics-acoustics.blogspot.com

الصوتيات ـ الأكوستيكا
مكتبة و ملتقى علم الأصوات
اللغة ـ السمع ـ الإدراك ـ النطق
www.facebook.com/groups/Phonetics.Acoustics

Table of contents

https://phonetics-acoustics.blogspot.com

CHAPTER 3:

EXPERIMENT METHODOLOGY

CHAPTER 4 :

RESULTS AND DISCUSSION

CHAPTER 5

https://phonetics-acoustics.blogspot.com

List of Tables

https://phonetics-acoustics.blogspot.com

https://phonetics-acoustics.blogspot.com

List of Figures

https://phonetics-acoustics.blogspot.com

List of Appendices

https://phonetics-acoustics.blogspot.com

Abstract

The subject of this thesis is to identify unknown speakers particularly from their speaking tempi represented in Speech Rate SR and Articulation Rate AR as temporal parameters.

The fundamental goal of this study, on the acoustical level, is to prove acoustically that every speaker has a significant speech rate SR and articulation rate AR through which the unknown speaker can be discriminated and to investigate which of them (SR or AR) could be more benefit for identifying unknown speakers and to what extent. Also, the present study is essentially concerned, on the perceptual level, with listeners' perceptual abilities in perceiving and differentiating different speaking tempi for identifying unknown speakers in order to utilize this exceptional ability in forensic speaker identification FSI.

The most important characteristic of the temporal aspects of speech (represented in speech rate SR and articulation rate AR) that they are not easily disguised or imitated by accent or fundamental frequency leveling; so they could be useful for identifying unknown speakers particularly in forensic phonetic field.

This present thesis is beginning with a brief theoretical and experimental review of other work in linguistic literature focusing on the temporal aspect of speech (speech rate SR and articulation rate AR) for identifying unknown speakers. And then the speech rate SR and articulation rate AR of ten unknown speakers / informants are calculated. The speakers were recorded while talking spontaneously for a radio program. Only 30 seconds of speech are cut for each speaker from the entire episode. After that 60 naïve listeners are asked to listen carefully to the 10 unknown informants in order to mark the fastest speaker and the slowest speaker depending only on their ears.

The present study focuses on the speech rate SR and articulation rate AR variations of colloquial Arabic speakers and aims to provide some useful acoustical and perceptual data in order to be used in forensic phonetic filed.

https://phonetics-acoustics.blogspot.com

CHAPTER 1

1. General Introduction

"A voice is more than just a string of sounds. Voices are inherently complex. They signal a great deal of information in addition to the intended linguistic message....."(Rose 2002)

The human voice carries on the speech signal which is a multidimensional and a very complex acoustic wave. These signals convey the information about the words or message being spoken in order to convey ***linguistic information*** (verbal) as well as ***non-linguistic information.*** **The non-linguistic information** reveals information about the speaker's identity, personality and individuality (***Extralinguistic component***[1]). It also reveals information about the situation, the inner state and the current health of the speaker, as well as their attitudinal or emotional state (***Paralinguistic component***[2]). That is, virtually any utterance conveys information on several levels (Hicks, 1979; Laver, 1994; Rose, 2002; and Trouvain, 2003).

The fundamental theory of forensic speaker identification relies primarily on that every voice is individually characteristic enough to distinguish itself through voice print analysis[3]. Many researchers support the theory that human voices are unique and could be used as a mean for identification. Besides, if everyone had the same voice, voices would not be used as discriminate evidence (Cain, Smrkovski &Wilson, 1990; Jessen, 2008; Rose, 2002; Nolan, 1997; Lindh, 2004; Mcpeek, 2013; and Owen, 2009).

Nowadays voice identification analysis has matured into a sophisticated identification technique. The comparison of human voices now focuses on every aspect of the words spoken; the words themselves, the way the words flow together, and the pauses between them. There are two methods of speaker identification as evidence which are commonly

[1]Extralinguistic information refers to aspects of the sound that are determined by the particular speaker’s vocal tract anatomy and physiology such as their vocal tract length or the volume of their nasal cavity. This component is uncontrolled and involuntary via speakers, and it is also called ***informative language*** because it is telling information about the informant / speaker itself.

[2]Paralinguistic information refers to habitual muscular settings that an individual adopts when they speak, for example; a speaker may habitually speak with slightly rounded lips, nasalization, or a low pitch range. This setting component is under a speaker’s control and voluntary. It is also called ***communicative language*** because it reflects the way speakers use language to communicate with others.

[3]The idea of the speaker identification is that every speaker has a unique voice pattern based on two factors: the first one is ***voice uniqueness***: including the uniqueness of the anatomical structure of every speaker. The second factor is ***the manner*** in which the articulators or muscles of speech are manipulated during speech: meaning the flexibility of the articulators which affected each others.

https://phonetics-acoustics.blogspot.com

applied: ***Naïve speaker identification*** (also called aural identification) which relays on human auditory perception and ***Technical speaker identification*** which uses acoustic analysis (through spectrograph) (Nolan, 1997 & 2001; and Rose, 2002). ***Naïve speaker identification*** involves the application of our natural abilities as human language users to the identification of a speaker (the term "naïve" means here the lack of specific training on the part of the person making the decision; for example: a witness to a crime may claim to identify a voice heard).***Technical speaker identification*** is defined by the employment of any trained skill or any technologically-supported procedure in the decision-making process when there is an incriminating recording of a suspect.

Both aural and spectrographic (visual) analyses are combined to form the conclusion about the identity of the speaker.Undoubtedly, the auditory method depends mainly on the naïve perceptual ability to recognize speakers' voices. Whereas; the acoustic analysis allows the phonetician expert to measure a number of parameters which reflect a person’s vocal organs; including fundamental frequency (pitch), formant frequencies(resonant), and the dimensions of the vocal tract and its pattern of movement. (Dellow, 2003; Lindh, 2009; Koenig, 1986; Eriksson, 2005; Lindh, 2004; Saeed, 2006; Rose, 2002; Owen, 2009; and Bonastre &others, 2003)

Your speech says very much about you. It can reveal your age, your health, your level of education, your regional dialect, and many other factors; even the location in which a recording is made. Thus, for a forensic phonetician expert, there's a wealth of information hidden in voices and this data is collected, observed, documented, compared and processed for forensic speaker identification FSI.

https://phonetics-acoustics.blogspot.com

- **<u>Thesis Structure</u>**

The present study is designed for identifying unknown speakers depending on their speaking tempo. This is be done through proving *acoustically* that each indiviual is differed from another according to the rate of speaking; and also through testifying *the listeners' perceptual abilities* in perceiving and differentiating different speaking tempi. Therefore; the entire structure of the presented thesis is based upon an *acoustical- perceptual line* (the acoustical measurements along with the perceptual abilities).

Nevertheless; the thesis is divided into two main parts: the first is a **theoretical part** that represents some history, definitions and problems; the second is an **empirical / experimental part** to illuminate some of the theoretical problems.

The theoretical part consists of two sections. ***The first section*** exposes a general introduction about the human voice usage and importance and the several levels of information it can reveal. Then, a clear distinction must be established between three terms; speaker recognition, identification and verification. After that, forensic speaker identification FSI would be exhibited as the most central aspect of forensic phonetics; staring with its beginning, definition, usage in different crimes, and the problems faced FSI as dogmatic evidence in courts. Then a general overview is presented about general automatic speaker recognition and its contributions in forensic speaker identification. At the end of this section, the main task of forensic speaker identification is explained relying on two main sources of variations of speakers.

The second section of the theoretical part deals with the literature of the most efficient acoustic parameters which are necessary for identifying unknown speakers; starting with the fundamental frequency and ending with speaking tempo which is the main concern of the presented thesis. After that, focusing on the temporal aspects of speech, the two main methods of measuring speaking tempo (speech rate SR and articulation rate AR) are identified in detail; in addition to demonstrating some factors that influence directly the rate of speech and its perception. Eventually, the perceptual abilities of the listeners is manifested; specially their perceptual abilities in speaking tempo identification.

The empirical / experimental part consists also of two sections: ***the first section*** involves the main experiment methodology and procedures; starting with gathering the data, the number of informants involved, the way of recordings, data analyses, the steps of analyses, the acoustical measurements, and finally the perceptual test procedures.

The second section of the experimental part contains all the results of the main experiment; the acoustical results and measurements with their statistical representations, for all of the involved informants, in addition to the perceptual results of the naïve listeners who joined in the experiment.

https://phonetics-acoustics.blogspot.com

1.1. Speaker Recognition, Identification and Verification: Some distinctions and definitions.

Speaker Recognition can be defined as any activity in which speech sample is attributed to a person based on acoustical and perceptual characteristics. Technically, it is used as a cover term for the wide variety of situations in which people are identified including all of the many different tasks of discriminating people depending on the sound of their voices.

Speaker Recognition has two subdivisions which are: a) Speaker Identification, and b) Speaker Verification.

Speaker Identification is considered as one of the subdivisions of speaker recognition which is the task of deciding and determining a given sample of speech (uttered by an unknown speaker who among many candidates said it) (N-Class division task, N→ the number of speakers). This process may be accomplished by humans alone, who compare a spoken exemplar with the voices of individuals. It may also be accomplished by computers alone, which are programmed to identify similarities in speech patterns, or it may alternatively be accomplished through a combination of humans and computers working together. (Furui 2008)

Speaker Identification is usually used in forensic situations where the circumstances are more difficult (the criminal sample may involve disguise, poor quality or distortion in transmission). Speaker identification has practical contributions especially in forensic speaker identification that arise out of a context of police work (forensic situations) or the presentation of evidence in court. Speaker Identification in the forensic context is usually about comparing voices. Probably the most common task involves the comparison of one or more samples of an offender's voice with one or more samples of a suspect's voice (Rose 2002).

Some tasks in speaker identification require strong interdisciplinary ties to speech technology, engineering or general acoustics that is the case with audio enhancement, where various forms of filtering techniques are used with the goal of improving the intelligibility of a low quality recording.

The second subdivision of Speaker Recognition is ***Speaker Verification*** which is considered as the task of accepting or rejecting, a given sample of speech (Furui 2008), whether a specified speaker said it or not; referred to speaker detection task (2-Class decision task). Speaker Verification also implies techniques by which a computer automatically compares the voice of the caller to a stored reference sample of the speech of the person whose identity is being claimed, Watt (2009).

https://phonetics-acoustics.blogspot.com

1.2. Forensic Speaker Identification FSI.

The term forensic phonetics has received official status at least since the foundation of the "International Association for Forensic Phonetics IAFP" in 1991 (Lindh 2009). *Forensic speaker identification FSI* is mainly concerned with solving problems related to identification of the unknown speaker in criminal investigation in which criminal intent occurs. FSI was developed by Bell Laboratories in the late 1940s for military intelligence purposes and the modern day forensic utilization of the technique did not start until the late (Cain, Smrkovski & Wilson 1990).

Forensic Speaker Identification FSI is the core of the scientific responsibilities of *forensic phonetics*[4] (Jessen 2008) whichis combined of both aural (listening) and spectrographic (instrumental) comparison of one or more known voices with an unknown voice for the purpose of identification or elimination (Cain, Smrkovski & Wilson 1990).

Jessen (2008) defines forensic phonetics as the application of the knowledge, theories and methods of general phonetics to practical tasks that arise out of a context of police work or the presentation of evidence in court; explaining that general phoneticians can make an important contribution to the forensic field like, running experiments with many speakers, investigating natural speech conditions, or presenting experimental results in a way that speaker variation can be seen, rather than pooling across speakers.

Rose (2002) defines forensic speaker identification FSI as the legal process as to whether two or more recordings of speech are from the same speaker, adding that, speaker identification in the forensic context is usually about comparing voices.

Also, Anna Salleh (2008) defines forensic speech as a science, which involves acoustic, phonetic and linguistic analysis of recordings that can be used to narrow down possible suspects and it can also be used to determine how well a suspect's voice matches a criminal's voice.

The use of forensic speaker identification FSI in criminal cases has a longer history than one might think (Butcher 2002). It has probably been used to identify suspects who were heard but not seen committing a crime for thousands of years. In more recent times there are many records of the use of speaker identification as evidence in courts; because of, the invention of telephones and recording equipment which opened new areas for forensic phonetics; and the analysis tools which are suitable for acoustic analysis of speech were also developed (Eriksson 2005).

[4]In other words; Forensic speaker identification FSI is the most central aspect of forensic phonetics (and acoustics).

https://phonetics-acoustics.blogspot.com

Therefore, Forensic speaker identification FSI has been used in a variety of criminal cases including; murder, rape, extortion, drug smuggling, blackmail threats and warnings, anonymous calls, ransom calls, kidnapping, wagering-gambling investigations, political corruption, money-laundering, tax evasion, burglary, bomb threats, terrorist activities and organized crime activities(Cain, Smrkovski & Wilson, 1990; Singh, Khan& Shree, 2012;Jessen, 2008;Nolan, McDougall, DeJong & Hudson, 2006; Lindh, 2004; Eriksson, 2005; and Nolan, 2005).

Watt (2009) and Jessen (2008) classify the majority of forensic speaker identification FSI work into two main types: ***speaker profiling*** and ***speaker comparison***.

Speaker profiling is carried out when no suspect has yet been identified, as for example when recordings of anonymous phone calls from kidnappers or bomb hoaxers are produced. Its purpose is to narrow down the population of possible suspects by identifying linguistic features associated with certain geographical and social groups, and any unusual pronunciations that may be attributable to exceptional anatomical or pathological characteristics.

While, ***Speaker comparison*** is based upon close comparison of two speech samples with a view to estimating the likelihood that the samples were produced by the same person. The expert’s task is to look for points of similarity and difference between the two voices (of the speaker in the ‘questioned’ or ‘disputed’ sample and that of a speaker whose identity is known).

As well as forensic speaker identification, *forensic phonetics* contains other applied areas such as ***general automatic speaker recognition GASR*** which is a speech-technological discipline that has been developed, and actually still developed, for purposes such as access control to sensitive information or to high-security buildings, devices or for other biometric means and applications (Jessen 2008).

Likewise, ***forensic automatic speaker recognition FASR*** is one of the applications of ***automatic speaker recognition ASR*** is in forensics. It's usually the case where a crime has been committed and the voice of the criminal needs to be verified from a recorded message; in order to prove that the suspect is the criminal; more specifically, that the voice of the criminal and the voice of the suspect are the same. Automatic speaker identification systems can be used as a method in forensic speaker identification dating back at least to the 1960s. But, it is observed that the error rates of machines are often more greater than those of humans. This is because of some problems that faced the approach of forensic automatic speaker identification (such as *within-speaker* and *between-speaker* variations) (Broeders 2001).Australia leads the world in the technology of

https://phonetics-acoustics.blogspot.com

automatic speaker recognition in 2001, but automatic speaker recognition is not yet able to separate out variation due to speaker differences from variation due to recording conditions (and it is doubtful whether it will ever be able to). The intervention of a phonetically and linguistically qualified human operator is required (Butcher 2002).

Resuming forensic speaker identification FSI, according to Rodman et al. (2002), FSI is the process of determining who spoke a recorded utterance; this process may be accomplished through three ways[5]: **the first way** is by humans alone, who compare a spoken exemplar with the voices of individuals. **The second way** is by computers alone, which are programmed to identify similarities in speech patterns. Or **the third way** is through a combination of humans and computers working in concert.

Obviously, a fully automated system for forensic speaker identification is not possible, because automatic speaker recognition and acoustic forensic analysis are completely different, especially in regard of the role of computers in both of them. That is in addition to two other respects; firstly, the degree to which some kind of human decision is involved. Secondly, the type of acoustic analysis implemented (Rose, 2002 and Mcpeek, 2013).

There are some problems that face and absolutely affect the strength of forensic speaker identification FSI as evidence. Eriksson (2005) states that the two main factors that may influence speaker identification accuracy are primarily ***sample duration*** and ***acoustic quality***.

Considering the influence of ***sample duration***, we may observe that in real life investigations samples may be very short, often just a few words or a phrase or two which means that sample duration is on the order of a few seconds. Eriksson (2005) adds that in an early study by Pollack et al. (1954) the authors observed that identification accuracy increased as sample size (for monosyllabic words) increased, but only up to about 1.2 seconds. For longer samples they claim that phonetic variation takes over as the most important factor. They conclude that; as they said:"we believe that the duration of the speech sample per se is relatively unimportant. The segments differed only in duration (25–2500 ms). For segments longer than about 75 ms, there was no increase in recognition rate as a function of duration".

[5]Broeders (2001) and also Mcpeek (2013) claim that there are two forms of forensic speaker identification FSI; ***the first*** and oldest form of speaker identification is of course speaker identification by ear witnesses which has been accepted as evidence by English courts since at least 1660, and it is very helpful, when it is the only way. ***The second*** major form comprises all forms of speaker identification by experts who working in the field of forensic speaker identification FSI using one of the three approaches (or the three types may combine in various ways): 1) Phonetic-acoustic approach, 2) Semi-automatic analytical acoustic approach, 3) Global automatic approach.

https://phonetics-acoustics.blogspot.com

With regard to ***acoustic quality***, this is a wide topic and can be thought of in many different ways. The most common quality problems in forensic phonetics are *background noise* and *bandwidth of recordings* or *transmissions*. By background noise we mean everything except the part of the speech signal, which contains information about the speaker we are interested in. It thus includes such diverse things as random noise as well as someone talking or a radio playing in the background.

Also Eriksson (2005) adds another factor, ***voice disguise***, which- to the extent that it is used- may be a serious problem for speaker identification. Besides ***performance of subjects*** referring to the performance of the human perceptual, storage, and retrieval mechanisms to decide whether the target voice is present in the line-up which allow a category of error. And ***style shifting*** which refers to speech variation *within* individual speakers depending on the speech context. (Nolan 2001), actually it will be a serious problem.

After all, many researches are needed to cover a significant area like forensic speaker identification FSI in order to enable it to affirm itself as evidence in court.

https://phonetics-acoustics.blogspot.com



1.3. Sources of variations of speakers.

The voice is not constant; the "voice" is defined as a combination of variables each of which has degrees of freedom. There are two main different factors that influence how people speak at any particular time and place. ***The first factor*** is out of the speaker's control, speech is the product of a combination of carefully controlled movements of the vocal organs which exhibit some considerable flexibility; such as the tongue and the lips (De Jong 2007). On the contrary, ***the second factor*** is depending on the speaker, speakers are perfectly able to adjust their speaking style or choice of vocabulary depending on the circumstances (for example; a casual style of speaking in conversations with friends and family or speaking more formal and choose their words carefully) (De Jong 2007).

All human voices are different due to the physical uniqueness of the vocal tract, the distinctive environmental influences in the learning process of speech development, and the unique development of neurological faculties which are responsible for the production of speech (Owen 2009).

A. <u>Organic vs. phonetic variability</u>

Thus, if individuals could be discriminated according to their voices, individuals' voices also must be crucially different. Indeed, they do (but the voice of the same speaker can also show considerable variation which must be discussed, see section **B**). These variations substantially arise from two principal factors (and also the whole theory of forensic speaker identification is based upon them); the first factor is *organic* and the second is *phonetic*.

The organic factor refers to the anatomical structure of the vocal tract from the larynx to the lips which is naturally attributed to genetic components (Mcpeek 2013). Definitely, every individual has nothing to do with the dimensions of his / her vocal tract, or of the thickness of the vocal cords. Those anatomical and physiological features of an individual are static to some extent; and of course, add some kind of uniqueness to the individual's voice (Watt 2009).

As already known; there are anatomical differences related to the sex (male or female); for example, men have longer vocal tracts than women. And also there are differences within the same sex related to the dimensions within the vocal tract (Beck, 1997; Nolan, 2005; and Jessen, 2008).

However further studies of identical twins by Nolan (1996) and Johnson (2000) indicate that despite those individuals having vocal tracts

https://phonetics-acoustics.blogspot.com

as nearly a like as two vocal tracts can conceivably be, as well as having had closely comparable parental input and social and educational backgrounds in the majority of cases, they still exhibit differences in speech production[6].

This is probably because of ***the phonetic factor*** which is substantially related to the manner and the style of the speaker. This factor depends mainly on the individual's speaking style and differs according to the speaker's emotional state (stress, sadness, or happiness), physical state(tiredness, fatigue, intoxication or having a cold), and to whom the speaker is speaking, orit even may differ according to related to different circumstances and conditions in which the speaker speaks. Moreover, the speaker has a degree of freedom involved in speech production, and also a huge number of possible combinations of segmental and prosodic features (Nolan, 2005; Rose, 2002; Dellwo, 2003; Krauss, Freyberg & Morsella, 2002; Jessen, 2008; and Watt, 2009).

Laver (1980) draws an explicit distinction between *organic factors* and *phonetic factors*. *The organic factors* studies the architecture and the anatomical structure of the speech organs resulted in the individual differences of the size, shape and length of the vocal organs. Whereas; *the phonetic factors* are concerned with the different ways in which an individual uses his or her voice and vocal apparatus which basically depend on the organic differences between individuals.

After all, there will almost always be organic differences (organic variation) which must have implications for an individual's speech production (phonetic variation) and if we ignore them, we will be ignoring a rich store of information which may help to explain some of the observed individual differences in speech production (Beck 1997).

B. ***Between*** **speaker vs.** ***within*** **speaker variability**

Everyone has a ‘voice’, but this is an oversimplification, firstly because a person’s voice is far from constant, and secondly because it is not yet clear how far the variations in a person’s speech cause their voice to overlap the ranges of other members of the speech community (Nolan 2001). So, *between* speaker and *within* speaker variability compose the second type of variation that contributes to make the voice of one speaker completely different and distinguishable from another. Generally, organic factors vs. phonetic factors are considered as the first source of variation; here a new distinction must be clarified between two other main kinds of

[6]Because, in recordings of speech, the effects of the organic differences are convolved with the effects of the speaker has learned (his/her personality characteristics). On the other hand, fingerprint is a direct trace of an organic difference only (uncontaminated by learning).

https://phonetics-acoustics.blogspot.com

variability: *between* speaker variability (which implies organic factors) and *within* speaker variability (which implies phonetic factors).

In view of the fact, as clarified in the previous section, different speakers have different sized vocal tracts giving rise to formants with absolutely different ranges of frequencies. Males and females speak with different fundamental frequencies F_0. Also, speakers from different regions and with different social backgrounds speak with different accents indicating that there is variability due to the speaker. (Ainsworth 1988, and Rose 2002)

Beck (1997) summarizes *within* speaker variability through determining three types of sources of individual variation:

1) *Life cycle changes*; these include the age-related changes in the vocal apparatus through the life span between birth and puberty.
2) *Genetic or environmental factors*; these include the variation in the precise nature and timing of the changes which occur during development and ageing and also coordinate with both endogenous and growth in some understood way to create the organic differences.
3) *Variation resulting from trauma or disease* (abnormal variation); includes the changes of the vocal apparatus which result from traumatic injury or disease.

Also Ainsworth (1988) confirms that even the same speaker speaking the same phrase on different occasions, he/she will speak it in different manners; different tempos for example. Moreover, speech will vary according to health and the emotional state of the speaker. Jessen (2008) adds that in a real situation, the acoustic environment will change and the background noise will contaminate the speech waveform. If you're tired, have a cold or speaking on a phone against traffic in the background you do all sorts of things to the voice, which make it phonetically very different from time to time (Salleh 2008).

In that way, an individual's voice characteristics can vary markedly. Although it is a general assumption that different speakers have different voices, it is crucial to understand that the voice of the same speaker will always vary as well. Rose (2002) said that, "It is a phonetic truism that no-one ever says the same thing in exactly the same way". And that is because of some of the acoustic differences that occur between repeats of the same word.

The features of speech and language can't be used as a marker of identity to identify one person and exclude all other people under normal circumstances, because people's voices overlap (Salleh 2008). Consequently, it seems an obvious and logical requirement of forensic

https://phonetics-acoustics.blogspot.com

speaker identification FSI, that the variation *between* speakers must be bigger than the variation *within* speaker. And after all, forensic speaker identification must be able to tell whether the inevitable differences between samples are more likely to be *within*-speaker differences or between-speaker differences, (Rose 2002). And of course the most powerful variations are definitely those that show a small amount of variation *within* a speaker and a large amount of variation *between* speakers.

Eventually, even within the bounds imposed by anatomy, physiology, dialect background, and so forth, individuals can still exercise a degree of choice over how they speak (Watt 2009). There will always be differences between speech samples, even if they come from the same speaker. These differences will always be audible, measurable and quantifiable, asserting that the main task of forensic speaker identification FSI is to find all the sources and types of variability in order to make a clear distinction for the correct evaluation.

https://phonetics-acoustics.blogspot.com

- **Aim of the work**

The explicit aim of this thesis is to clarify acoustically and perceptually the role of the speaking tempo (SR and AR) in identifying unknown speakers.

Consequently, this aim will contribute in giving an overview of all the following levels:

1- Discussing whether Speech Rate SR or Articulation Rate AR is more useful and more helpful in identifying unknown speaker.

2- Exploring the extent to which the listeners' ears can realize the difference between slow and fast speaking tempi.

3- Finding out the most acoustic cues that listeners, perceptually, utilize to realize the tempo of speaking.

https://phonetics-acoustics.blogspot.com

https://phonetics-acoustics.blogspot.com

CHAPTER 2

2.1. The most efficient acoustic parameters for identifying unknown speakers.

There are four main phonetic parameters depending on the speaker through them the speaker can be discriminated and identified (in forensic situation); those parameters are:

A. The Fundamental Frequency F_0.
B. The formants frequencies of the vowels.
C. The resonance of the nasal consonants.
D. Speaking Tempo.

Considering that all the coming parameters together, which are responsible for voice individuality, can give a complex profile of the speaker based on natural data and extracted from his/her speech signal. Keep in mind, there is no ideal parameter, but the most important criterion is for a parameter to show a high ratio of between-speaker to *within*-speaker variation (Rose 2002). Absolutely there are many different parameters that can be used to compare speech samples forensically.

A. <u>Fundamental frequency F_0</u>

The first most commonly used measurement in the identification and discrimination between speakers is average fundamental frequency F_0 and the glottal waveform, because it is different from one speaker to another and it is referred to *the source*, which is what the listener perceives as the pitch of the voice (Battaner et al., 2003; Lindh, 2006; and Rose, 2002). Cicres (2011) examines the preliminary parameters which discriminate speakers in forensic phonetic field. The experiment includes 150 hesitation sounds from 15 speakers to study the acoustic parameters of the voice quality. The results indicate that fundamental frequency F_0 showed significant differences between speakers. Also, Mikos & Gelfer (2001) studies the relative contribution of speaking fundamental frequency and formant frequencies to gender identification of biological males and biological females based on isolated vowels to determine if fundamental frequency or formant frequencies of vocal tract resonances play an important role in identifying speaker's gender. The experiment included 10 males, and 10 females who served as speaker subjects. Each speaker was recorded sustaining the vowels / i /, / u /, and / a / for five seconds, at comfortable conversational levels of pitch and loudness. The results showed that pitch was the dominant cue and it is accounted for about 80% and formant frequencies resonances about 20% of correct gender identifications.

Fundamental frequency $F_0$is influenced by paralinguistic and other types of variation, and indeed it is one of the main ways of conveying both grammatical and emotional meaning in speech. Busso et al. (2009) analyzes the statistics of the pitch contour through detecting pitch features derived from emotional speech samples compared with the ones derived from neutral speech. The results indicate that gross pitch contour statistics such as mean, maximum, minimum, and range are more emotionally prominent than features describing the pitch shape. Also, analyzing the pitch statistics at the utterance level is found to be more accurate and robust than analyzing the pitch statistics for shorter speech regions (voiced segments).

Each speaker has a particular range of fundamental frequency F_0 which he/she habitually uses and within which he/she feels most comfortable and this is an important measure for forensic purposes. The average speaking fundamental frequency F_0 for an adult male is 113 Hz, in spontaneous speech and the corresponding average for females is 225 Hz. (Butcher, 2002; Hicks, 1979; Rosen, 1992; and Lindh, 2009). Morris (2011) investigates the contribution of F_0 in identifying speaker's gender from speech sounds relying on speakers' voice fundamental frequency (F0). The experiment includes 200synthesized vowels with a range of F0-formant combinations. The synthesized vowels were presented to 28 native Cantonese-speaking listeners to judge the perceived speakers' gender for each of the synthesized stimuli. The results revealed that F0 was the primary cue for speakers' gender and when F0 was below 162.01 Hz or above 204.97 Hz, listeners reliably and correctly identified the speakers as male or female, respectively.

Fundamental frequency F_0 has been used for a long time in speaker identification, because F_0 has been shown to be a successful forensic phonetic parameter. But to be able to study differences *between* speakers; the duration of the samples should be more than 60 seconds (Lindh 2006). On the contrary, Rose (2002) emphasizes the importance of fundamental frequency (F_0) especially in forensic speaker identification FSI; stating that it can be extracted with relative ease from poor-quality recordings and it is not adversely affected by telephone transmission, unlike F1 and the higher formants. However, any changes in health that affect the size or shape or organic state of the vocal tract, or its motor control, will alter its acoustic output.

Perceptually, voices differing in average F_0 can be described on a scale between 'high-pitched' and 'low-pitched' (Ainsworth 1988).Therefore, some phonetic parameters are better captured acoustically than auditorily; and the most classic one of these in speaker identification is the average fundamental frequency F_0 of the voice (Jessen 2008).

https://phonetics-acoustics.blogspot.com

B. Formant frequencies of the vowels

The second most commonly used measurement is the formant frequencies. When a speaker pronounces a vowel sound, the frequencies at which there is maximum energy are called resonant frequencies; where the air in the supra-laryngeal vocal tract is vibrating with maximum amplitude at these frequencies. In acoustic phonetics, these are known as formants. Formants are very commonly used in forensic speaker identification, both in experimental work and case-work (Rose 2002); because formant frequencies of vowels are referred to a set of characteristics related to *the filter*. La Riviere (1975) examines experimentally the relative contributions of fundamental frequency and formant frequencies in aural speaker identification. Speech samples were four voiced, whispered, and low-pass filtered isolated vowels produced by eight male speakers. Twelve listeners were asked to make speaker identification choices from these stimuli. Utterances were subjected to acoustic analyses. The results indicate that both fundamental frequency and formant frequencies contribute, approximately equally, to speaker identification judgments.

One of the anatomic foundations of formant frequencies is their dependence on vocal tract length. So that, mean length of the vocal tract of the speaker can be obtained from the first and second formant frequencies (F_1 and F_2). Also F_3 mean value (the third formant frequency) is particularly useful because it does not change considerably from a vowel to another and gives more accurate information than F_1 and F_2 (Battaner et al, 2003; Stevens, 1998; Rose, 2002; and Ainsworth, 1988).

Lindh (2006) states that the first four formants for a selected vowel; have an important role to examine their values in identifying speakers. In addition to the role of the higher formants which is not necessary for speaker identification, but in some cases they provide some useful information. McDougall & Nolan (2007) studies the discrimination of speakers using the vowel formants of /uː/ in British English. The experiment includes six repetitions for each speaker of the vowel /uː/ in hVd context. The results indicated that F_2 analyses consistently provide higher levels of classification and exhibit greater *between* speakers' variations than F_1 analyses. Also, Loakes (2006) investigates the spontaneous conversational speech patterns of twins who have had similar educational and environmental influences. Comparisons of twins' formant frequencies (F_1 –F_4) showed that F_3 is the most speaker-specific formant frequency.

Whereas, Rose (2002) states that higher formant frequencies (F_3 and up) seem to have a particular importance especially for forensic speaker identification FSI; because they reflect individual characteristics more than those in the lower frequencies (F_3 and below). F_3 is a good indicator

https://phonetics-acoustics.blogspot.com

of a speaker's vocal tract length, and F_4 and F_5 are 'indicative of a [speaker's] voice quality'. In addition; higher formants are high speaker-specificity because they are less free to directly reflect individual anatomy and reflecting the resonances of relatively fixed smaller cavities in the vocal tract, for example the larynx tube, which are assumed to be relatively unaffected by the gross configuration changes of the vocal tract involved in the production of different vowels. This would contribute to a desirable smaller degree of *within*-speaker variation. Wilson et al. (2006) studies set of acoustic parameters that attempt to capture speaker-specific characteristics. The experiments include telephone speeches. The population size of each experiment was varied from 50 to 250 speakers and they were either all male speakers or all female speakers. Their results showed that F_1 and F_2 vary considerably due to the vowel being articulated, whereas F_3 and F_4 change very little. The frequencies of the formants during sonorant sounds provide information about the length and shape of the vocal tract. The vocal tract length of a speaker can usually be characterized by F_3. However, they have found that F_3 and the higher formants may be indicative of vocal tract shape during sonorant consonants. For example, the American English /r/ sound can be produced with a variety of vocal tract configurations that all give a very similar acoustic profile for F_1-F_3. However, it appears that the higher formants, in particular F_4 and F_5, may be acoustic signatures of the differences in vocal tract shape.

After all, speakers cannot change the length of the vocal tract which limits the range of formant frequencies (Dellow 2003). Therefore, Butcher (2002) points out that about 50% of the unknown speaker's vowels fall within the corresponding ellipse of the suspect speaker. Based on this data alone, there would have to be considerable doubt that the speakers are the same. So, data from vowel formant frequencies alone is not considered to be enough in identifying unknown speakers and it must be combined with other parameters.

C. The resonance of the nasal consonants

The third parameter is the resonance of the nasal consonants. Some speakers are known to characteristically speak with a "nasal twang" (which is extra-linguistic information). Speakers who have them are usually not aware that they do, and even if they are, still do not have much control over them because it's related to anatomical foundations. The resonant frequencies associated with the nasal cavities, which are the most fixed structures in the vocal tract, and they are influenced by the degree of the soft palate opening and by different vowels (Rose 2002).

Nasal consonants like [m, n or ŋ] are forensically relevant because their acoustic characteristics are assumed to be among the strongest automatic speaker identification parameters. This can be relevant in voice

https://phonetics-acoustics.blogspot.com

disguise, where a criminal might concentrate on one type of disguise (such as falsetto voice), but be unable to control other aspects of his speech, such as nasalization and glottalization. Also filled pauses can differ with respect to presence or absence of a nasal consonant. They can also differ in other respects, such as their duration or frequency of occurrence in a given time interval, which is better captured with measurements in the acoustic signal (Rose 2002).

Perceptually, Amino & Arai (2009) studies the effects of the oral and nasal consonants and vowels on perceptual speaker identification of familiar and unknown speakers in Japanese. The experiments include four analysis intervals through six monosyllables uttered by ten speakers: /ta/ /da//ma/ /na/ /sa/ and /za/. The results shows that nasal sounds contain more individuality than oral sounds, where, the stimuli including a nasal sound were effective for both familiar and unknown speaker identification by listening. Moreover; according to the place of articulation, coronal nasals /n/ and/ɳ/ were better than labial /m/. Also, the vowel following a nasal consonant is necessarily nasalized to some extent, and nasalized vowels are predicted to contain more individuality than non-nasal vowels. Furthermore, co-articulations to the following vowel are also greater in /m/ than in /n/.

Nasal resonance can be an effective cue to speaker identification, because of the relative rigidity and stability of the nasal cavity compared to the fundamental frequency or other phonation parameters. These properties ensures a low *within*-speaker variation in the acoustic features associated with the cavity's acoustic resonances, and its internal structure and dimensions are complicated enough to contribute to relatively high between-speaker variation. Therefore, the resonance of the nasal consonants has a remarkable importance in forensic speaker identification FSI. Whereas; on the other hand, nasal resonance properties are influenced seriously by head-cold, hay fever and other nasal diseases as the transmission through the nasal tract may be affected.

D. Speaking Tempo

The fourth parameter is the speaking tempo which reveals the temporal aspects of the speech. Tempo of speaking is the subject of the present thesis and will be illustrated in details in the next section (2.2.Tempo of Speaking).

Keep in mind that there are various terms used to denote the tempo of speaking such as speech rate, rate-of-speech (ROS), rate of speech production, speed of talking, talking rate, reading rate (for read speech), speaking tempo or simply tempo. These terms are used here as synonyms and most of the time in this thesis the expression speaking tempo is used.

https://phonetics-acoustics.blogspot.com

2.2. Tempo of Speaking

Tempo of speaking is a multidimensional phenomenon which comprises various aspects (Kohler 1986); and it is one of the supra-segmental features which realize or express prosodic contrasts. Tempo of speaking is considered as a non-linguistic cue that signals paralinguistic information as well as extralinguistic information(Laver 1994).Those non-linguistic factors have significant importance particularly in forensic speaker identification FSI; because they are carrying the individual-identifying information about the speaker. For more illustration, although tempo is a prosodic property, tempo is not a genuine linguistic property, because it is not used to distinguish linguistic structure. Moreover, linguistic structures require units which can be described as invariant and distinctive. While on the contrary, there is no direct linguistic property or contrast that can be attached to speech tempo. In addition, tempo neither bears any meaning nor does it differentiate any meaning by itself. However, tempo can have a strong effect on the realization of linguistic structures, for example; the sentence "John loves Mary." spoken slowly means exactly the same as the same sentence spoken a bit faster. But undoubtedly, tempo can reveal some paralinguistic and extralinguistic information about the speaker's overall level of pitch, or emotional state. Real speech varies in its speed of utterance; speaking is as every other form of motion, takes place by continuously moving parts of the body over time, consequently there is always a given tempo of speech. On the production side, speakers are able to change their rate of speech if they deliberately intend to do. So as to, on the speech perception side, listeners have an idea whether a given portion of speech was fast or slow relative to an expected normal tempo. But usually changing the rate of speech happens unconsciously; because the speaker can't control the way, rather than the rate he/she is speaking in (Trouvain, 2003; and Rose, 2002).

Long time ago, many researchers assure that temporal parameters are very important for speaker identification particularly in forensic situation (Battaner & others 2003). Doherty (1975) examines the effectiveness of certain acoustic and temporal properties of the speech signal in the determination of a speaker's identity from his voice alone by extracting speaking fundamental frequency (SFF), and speaking tempo (ST) separately from a speaker's vocal output. This provides sufficient bases to make judgments of a speaker's identity. The results of the experiment conclude that speaking tempo seems to have a greater capacity to determine the identity of speakers who are attempting a disguise. It's appeared that speakers do not attend to that temporal factor while they are attempting to disguise their voice. Sequentially, speaking tempo is a very important parameter in forensic speaker identification FSI because

https://phonetics-acoustics.blogspot.com

different speakers can habitually have different speaking tempi. Also, Kohler (1986) states that the speaker identification process depending on temporal parameters of speech, in the sense, shifts to higher ratios for slower speaking tempo and to lower ratios for faster speaking tempo.

Furthermore, Demenko (2000) states that tempo of speaking has a significant importance in Forensic Speaker Identification FSI because it is:

1. Carrying the individual-identifying information about the speaker.
2. Affected by the individuals variations in speaking.
3. Not affected by the frequency characteristics of the transmission systems and at the level at which the speaker talks.
4. Not easy to imitate or disguise.
5. Not controlled by the speaker.

Individual speakers can differ substantially in their typical speaking tempo. This can easily be shown by looking at a database containing the same text read by many different speakers. Furthermore, changes in speaking tempo happen all the time, all day long, there are numerous situations, conditions and circumstances in which these changes take place. Of course, the speed of speech is not constant, and it is a remarkable fact that we can both produce and understand speech at very different tempi. Accordingly, an important distinction must be drawn between the subjective tempo in production and perception on one hand, and a measured objective tempo on the other hand. When a speaker intends to speak "fast", the resulting speech will not necessarily be categorized as "fast" on the basis of a physical measurement (e.g. in syllables per second). In other words, the durations of the various productions as objective measurements do not necessarily mirror the intended tempo as a subjectively produced. Therefore, Trouvain (2003) distinguishes between three types of speaking tempo:

1) ***The subjective***: intended tempo of the speaker's speech production.
2) ***The objective***: measured tempo reflected by durational correlates of linguistic units.
3) ***The subjective-perceived tempo***: reveals speaking tempo that the listeners perceive.

- **Speech rate SR and articulation rate AR**

Speaking tempo can be exhibited by two methods, one is ***speech rate*** / ***speaking rate*** / ***syllable rate*** (all terms can be abbreviated to **SR**), and another is ***articulation rate*** (**AR**). Both of SR and AR can be defined as "*the number of output units per unit of time*", (that unit might be phone, syllable, token or word. e.g., syllables per second). The biggest difference between SR and AR is that the SR includes pause intervals but the AR does not. In other words, the usual distinction is that speaking tempo can

https://phonetics-acoustics.blogspot.com

either be defined as articulation rate or as speech rate. Articulation rate as a net rate refers to phases of articulation excluding pauses. Speech rate as a gross rate refers to the entire speaking phase including pauses (Gold, 2012; Koreman, 2006; Rose, 2002; Trouvain, 2003; Fashal, 1991; Kohler, 1986; and Hicks, 1979).On the other hand, Laver (1994: P. 541) pointed out that there are several relations between articulation rate and speaking rate depending on the continuity of the speech. Illustrating that; a fast articulation rate could be combined with a fast overall speaking rate if the speech is fluent. And also a fast articulation rate could be combined with a slow overall speaking rate if the speech is interrupted with long pauses.

Miller & others (1984) defines speech rate SR as a complex variable and it is composed of both the rate at which the speech itself is produced (articulation rate AR) and the number and duration of pauses in the utterance (pause rate). The paper is proposed that the overall changes in speaking rate SR that occur *within* and *between* speakers are largely due to changes in the pause rate while the articulation rate AR is remaining relatively stable across different conditions and conversations. Whereas, Dankovicova (1999), while investigating articulation rate variation within the intonation phrase in Czech and English, discusses that AR does not vary in an arbitrary way but its variation is structured. It is structured both in terms of having a domain (the intonation phrase), and in terms of a recurrent pattern of slowing down throughout the phrase. Therefore, since people tend to pause during speech, speech rate SR is usually lower than articulation rate AR. However, the articulation rate AR showed some promise as a forensic-phonetic parameter.

If the difference between speech rate SR and the articulation rate AR can be so troubling due to a pause, then the definition of a pause is crucial to determine speech tempo. Pauses of speech can be divided into two types: filled pauses and silent / unfilled pauses (see section 2.3.3 for more illustrations). Consequently, Laver (1994) defines the articulation rate (AR) as the rate at which a given utterance is produced excluding only the silent pauses, and he supposes that the articulation rate AR must include all the audible / vocalized speech material (such as *um*, *mm* and the prolonged syllables) with the utterance because that materials make up the linguistic message of the utterance with any filled pauses. Accordingly, Rose (2002) distinguishes two different measures of tempo, although their precise definition varies a little. The first measure is the articulation rate (AR), and it is defined as "the speed at which an individual utterance is produced". AR is calculated basically by the number of syllables in an utterance divided by the utterance duration excluding only silent pauses. The second measure is the speaking speech rate SR, and is defined as the overall tempo of performance, and includes all pauses. SR calculates basically the number of syllables in a set of utterances divided by the overall utterance duration including pauses.

https://phonetics-acoustics.blogspot.com

2.3. Factors that directly influence the rate of speech and its perception.

Measuring speech tempo contains various sources of confusion. There are many phenomena observable in speeded up and slowed down speech. There are lots of parameters that influence the rate of speech as well as its perception (see figure 1). Speakers vary in certain aspects of their speech, but these factors particularly; speakers cannot control in order to change their rate of speaking. In addition, the listeners use these factors for making tempo judgments about the speech sample.

Moreover, if we ask listeners to judge which of two recordings of the same text they think is faster, the choice does not necessarily fall onto the production with the shorter duration. For this judgment, other factors could play a role such as dysfluencies, assimilations, number and duration of pauses etc. Moreover, there is also an influence of fundamental frequency on perceived rate. Which means; that the subjective impression of speech tempo does not exactly match the objective measurement (Trouvain 2003).

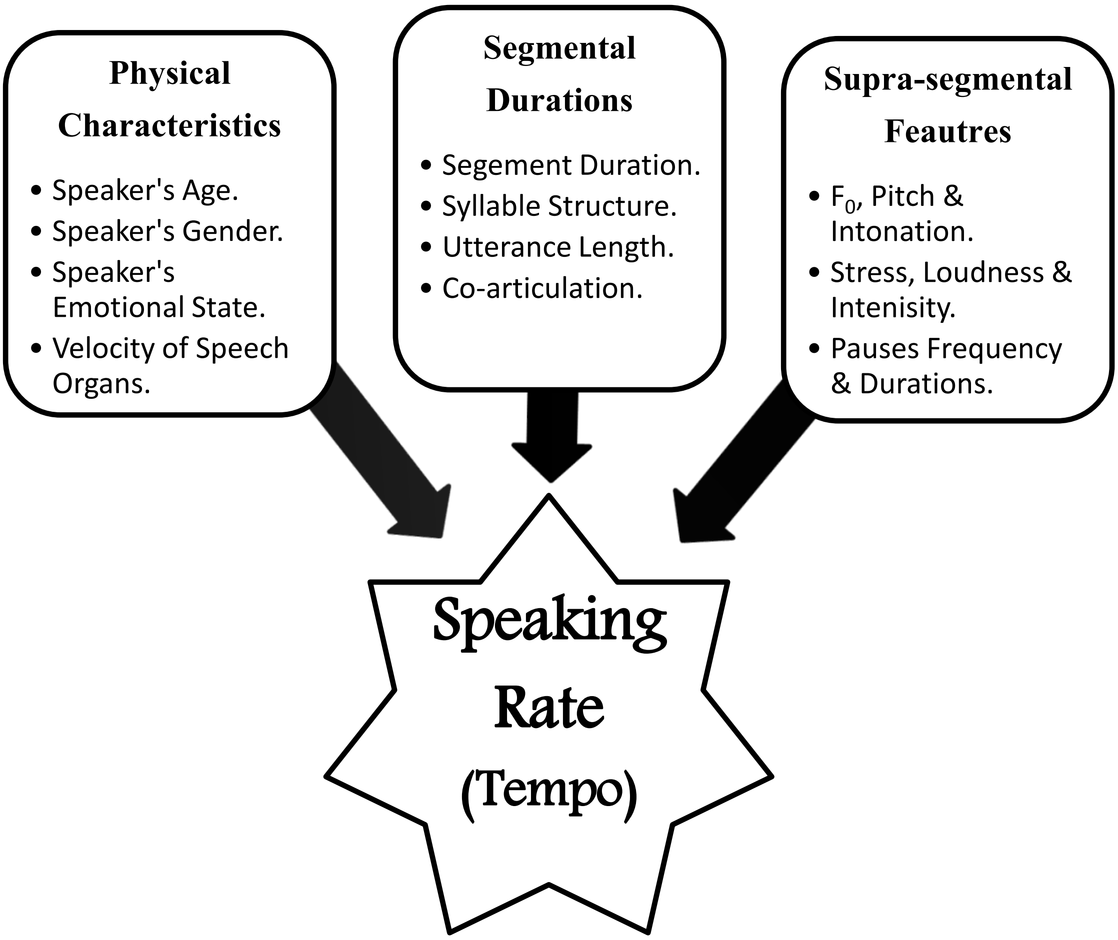


Figure 1: showing factors that directly influence speaking tempo and its perception.

https://phonetics-acoustics.blogspot.com

2.3.1. Tempo and Physical Characteristics.

Speaking tempo is interconnected with the physical and psychological characteristics of the speakers; such as: speaker's age, gender, emotional state and the degree of the velocity and movements of the speaking articulators.

Regarding ***the age of the speaker***, Starkweather (1987)[7] assumes that the ability of speaking fast increases with age so as to a number of other skills like skills to move rapidly. Supporting his assumption, those skills originated in increasing the syntactic, semantic, and phonological knowledge which requires more complex and longer speech utterances. This assumption is adopted experimentally by Kowal et al. (1975)[8], who showed that the rate of spontaneous speech actually increases with age; between 6 and 14 years old, the rate of speech increases roughly from two to four syllables per second. Also Haselager et al. (1991)[9] investigates articulation rate skills of Dutch-speaking boys and girls in four age groups between 5 and 11 years through repeating the same syllable as fast as possible as well as spontaneous speech elicited in interviews. For both speaking modes the syllabic rate varied with the age group and their results indicate that the younger the group the slower the articulation rate, which absolutely assure the assumption that the articulation rate increases during maturation.

On the other phase, the effect of the speaker's age seems to apply not only to the developmental phase of speakers (during maturation), but also Malécot et al (1972)[10] find in their study on French spontaneous speech that older adults speak slower than younger adults. Furthermore, syllabic rate drops progressively by about half a syllable per second overall from 5.95 syll. /sec to 5.52 syll. /sec, for young (starting at 20 years) to older speakers (up to 69 years). Moreover, Schwaba&Avanzi (2015) while examining the articulation rate in read speech and in spontaneous speech in seven variants of French (Paris, Lyon, Tournai, Liège, Geneva, Neuchâtel, and Nyon). They observe that speaker's age influence the articulation rate. Articulation rate decreases with older speakers more strongly in the Swiss and Lyon variants than in Paris and Belgian variants.

https://phonetics-acoustics.blogspot.com

[7]Cited in:Slis, I.H., Haselager, G.J.T. & Rietveld, A. C. M. (1991) Development of Speech Rate inDutch Children. *Clinical Linguistics & Phonetics.*

[8]Cited in:Slis, I.H., Haselager, G.J.T. & Rietveld, A. C. M. (1991) Development of Speech Rate in Dutch Children. *Clinical Linguistics & Phonetics.*

[9]Cited in:Trouvain, J. (2003) *Tempo variation in speech production, implications for speech synthesis.* PH.D dissertation.

[10]Cited in Trouvain, J. (2003) *Tempo variation in speech production, implications for speech synthesis*. PH.D dissertation.

With regard to ***the gender of the speaker***, some studies indicate that men actually speak somewhat faster than women do. Whiteside & Hodgson (2000), in their acoustical comparative sex studies from a developmental perspective, found a significant effect of gender on the speech tempo; by means, females articulate slower than males; except for the six years old. Also, Jacewicz et al. (2009) proves this experimentally during their investigations on speech tempo differentiations across dialect, age and gender between northern and southern speakers of American English in reading and informal talks. The results indicate that the effects of gender were smaller and less consistent; however, men generally spoke slightly faster than women did. Moreover, Schwaba & Avanzi (2015) examine the articulation rate in read speech and in spontaneous speech in seven variants of French (Paris, Lyon, Tournai, Liège, Geneva, Neuchâtel, and Nyon). The results showed that males had faster articulation rates than females (188 ms and 198 ms, respectively), and also the differences between males and females vary as a function of the regional variant.

However, Malécot et al. (1972) found no significant differences between the sexes in terms of syllable rate, but there were differences in terms of utterance length for French language. O'Neill (2008), as well, examines speaker's gender differences in speaking rates in read sentences and spontaneous speech. The results indicate that the differences in articulation rate as a function of gender were very small, whereas; as a general tendency, men spoke faster than women in read sentences (3.47 syll. /s vs. 3.33 syll. /s) and in spontaneous talks (5.23 syll. /s vs. 5.04 syll. /s) but the differences were very modest. For read sentences, the articulation rate for men was 4% faster and for spontaneous talks, it was 3.6%. So, the results of either articulation rate in read sentence or spontaneous speech showed no significant main effect of speaker gender.

Whereas, the findings of Deese (1984) for English language contradict the aforementioned studies, his results showed that women spoke faster (5.82 syll. /s) than men (5.48 syll. /s) but this difference was not statistically validated. Also, Clopper & Smiljanic (2011), while studying the effects of gender and regional dialect on tempo in American English, their experiment included five males and five females who read two conversational passages. The results showed significant effect of the speaker's gender on the distribution of pitch accents and pauses, and they observed that male speakers paused more frequently than female, which give an impression of slow speaking rate in male speakers' speech.

Thus, the influence of gender on speaking tempo is remaining an open question in debate because there is a lack in studies of social sources of

https://phonetics-acoustics.blogspot.com

variation that are based on sex / gender differences on speech tempo, so it can be particularly problematic.

Speaking rate is also correlated with ***the speaker's emotional state***. In such a manner, a fast speaking tempo may express irritation or urgency, whereas a slow speaking tempo may express hesitancy, doubt or boredom (Laver 1994).

Emotions could be divided into two basic groups according to temporal measurements. *Anger, rage, fear,* and also *happiness* form one group which is generally marked by an increased tempo; and characterized by rapid speaking rate and shortened durations of phonation and pauses, but did not differ significantly from each other. Whereas, *boredom, sadness, sorrow, grief, contempt,* and *disgust* form another group which is characterized by a slow rate of speech tempo, but each emotion of this group could be differentiated from the other emotions, for example; the slow rate for grief results almost entirely from prolonged pauses, particularly those between phrases, whereas the slow rate for contempt is because of increasing the duration of both speech segments and pauses (Hicks1979).

Lee et al. (2005) investigates the speech articulation in association with four different emotional speech production (neutral, anger, sadness and happiness), through attaching three sensors to three major speech articulators (the tongue tip, the lower maxilla for jaw movements and the lower lip). The results show that utterance durations become longer when subjects spoke in emotional mood. In addition to higher pitch and wider pitch, distributions are observed specially with happy and anger emotions. Lee et al. (2005) summarizes that differences in movement ranges and velocities among the four emotions are clearly observed which indicate that all emotions are effectively discriminated in temporal articulatory domain.

When human articulate faster; different mechanisms of articulatory organization may come into play. The effects of speaking rate variation appear to have particularly complex influences on articulation. This takes us to another important factor that is: the ***velocity of the speech organs***.

According to Berrya & Weismer (2013) there are four primary measures of interest that can be identified for the velocity of the speech organs: 1) articulatory velocity (speed); 2) movement extent (distance); 3) relative timing (phasing); and 4) velocity profiles.

In speech, the articulators are our bodies moving in time and space; timing rules determine the movements of the various articulators and their temporal interrelations (Lisker 1989). However, with respect to the articulators there is a lack of homogeneity; some articulators move inherently faster than others (for example; the tip of the tongue is able to

https://phonetics-acoustics.blogspot.com

execute many more movements in a given time compared to the velum). Moreover, the articulators neither move all at the same time nor at the same extent (Trouvain 2003).

Masaki et al. (1986) discusses that the jaw opening for each vowel at the fast speaking rate was less than that at slow speaking rate, explaining that by undershoot due to the shortness of the duration. In addition, the variation of the jaw opening for each vowel is depending on the vowel's position in the word. Also Fletcher (1985) confirms that, in fast speaking tempo; there is less pressure on the articulators (speech organs), resulting in undershoot or executing the gestures which is necessary to produce all of the segments.

Moreover, Ostry & Munhall (1985) study the tongue movements during the production of CV sequences which contain all the combinations of the back vowels and velar consonants, produced in fast and slow speaking rates. Their results indicate that fast speaking rate is characterized by a large ratio of maximum velocity to movement amplitude through decreasing the tongue movement duration; and conversely, slow speaking rate is characterized by small ratio and increasing tongue movement duration.

Also, McClean & Clay (1995) examine the activation characteristics of lip-muscle single motor units in relation to speech rate; through describing how lip-muscle motor vary with different speech rates. Their experiment includes three male adults; their task is to repeat CV syllables at four different rates. The results showed that lip-closing velocities tended to increase with increasing the speaking rate; whereas, lip-opening velocities remained relatively constant.

Moreover, Lee et al. (2005) while investigating the speech articulation in relation with different emotions; observes that angry and happy speech showed the largest velocity of the tip of the tongue, and very small tongue tip velocity for sadness speech emotion.

Thus, the articulators' velocity seems to vary for individual speakers, because of that, the articulators' velocity of a speaker has a significant importance as a factor that affects the overall speaking tempo especially in forensic speaker identification FSI field.

https://phonetics-acoustics.blogspot.com

2.3.2.Tempo and Segmental Durations

Segments durations cannot be easily ignored because of their absolute effect on the total duration of the perceived utterance. A change in the articulation rate does, of course, imply a change in the temporal extension of speech stretches corresponding to linguistic elements such as syllables and sound segments. Segments reveal different behavior in terms of compressing and expanding their durations according to their sound class.

Fletcher (1985) examines some effects of different tempi in French. Her experiment includes native French speakers who were asked to read a passage at increasing tempo from slow to normal to fast. Her results showed that the involved speakers compress the segments durations (consonants and vowels) and some segment elision and syllable loss was occurring due to fast speaking tempo. These results indicate that segment compression is a most important source in increasing the speaking tempo. Moreover; segment compression and expansion are considered as an obvious universal strategy for speakers to modify their speaking tempo. Fletcher explains that , in fast speaking tempo, consonant and vowel articulations following each other in quicker succession may cause the reduction in duration and quality of the vowels; particularly if unstressed, so as to consonants.

In addition, Dellwo, Leemann & Kolly (2012), studied the speech timing in the Zurich variety of Swiss German during reading conversational speech materials. Their results revealed that there is temporal variability on the segmental level even in spontaneous speaking. More specifically, the vocalic durational information is a more vital source of durational information between speakers than a consonantal one because consonants may not have enough articulatory flexibility in their production in contrast to vowels. Also, Ghazali et al. (2002) in their experimental investigation about speech rhythm variation in Arabic dialects showed that the reduction of vowels in the Western dialects (like Moroccan dialects, Algerians and North African Arabic dialects) mainly affect the speaking tempo. Whereas, longer vowels in Eastern dialects (like Tunisian, Syrian and Jordanian dialects) are considered as a significant factor in differing speaking tempi, thus vowels expand and compress more than consonants. As a matter of fact, slow articulation means primarily vowel lengthening, and fast articulation means primarily vowel shortening. However, there are many factors responsible for segment durations (Trouvain 2003).

Gay (1978) studies the effect of different speaking rates on vowel formant movement. His experiment includes four speakers who produced different syllables at fast and slow rates. The results showed that the

https://phonetics-acoustics.blogspot.com

formant movement of the vowel began earlier in fast speaking rate and also that increasing the rate of speaking leads primarily to decreasing the duration of the vowels.

Also, Van Son& Pols (1992) investigate the effect of normal and fast speaking rates on the dynamic structure of Dutch vowels formants in reading. Their experiment includes a single professional Dutch speaker who reads a meaningful text of 844 words in two different rates: fast and normal. The results of the experiment indicate that fast rate tokens are shorter than the normal rate tokens; and also the vowels / a / & / o / show a shortening of around 20 % at fast speaking rate; whereas, short vowels did not differ in length between fast and normal rates.

Furthermore, Hirata & Tsukada (2004) examines how the vowel formants of Japanese vowels (short and long) vary as a function of different speaking rates. Their experiment includes two male native speakers of Japanese asked to produce ten non-words including five short and five long Japanese vowels. Their results showed that the vowels / e / & / o / were mostly affected by the changes of speaking rate and the vowel / i / was the least affected.

Moreover, Magen & Blumstein (1993) studies the effect of speaking rate on short and long vowels durations in Korean. Their results show that short and long vowels were similarly influenced by the different speaking rates, and also the durations of the short vowels spoken at a slow rate were always longer than the short vowels spoken at fast rate, and that short vowels spoken at a slow rate almost always overlap in duration with the long vowels spoken at fast rate.

On the perceptual level; Halle (1998) studies the subjective speech rate, attempting to examine the effect of intra-syllabic structure on perceived tempo. The results clarify that the subjective articulation rate is affected by the vowel duration and the listeners are highly sensitive to vowel nucleus durations which indicate that the intra-syllabic structure of speech may modify the overall speaking tempo.

Jessen (2008) emphasizes that the measurements of the sounds durations (segments' durations) is a promising area in forensic speaker identification that pertains to speech tempo. Based on the results of Allen et al. (2003) who address the question whether known speaker differences in Voice Onset Time VOT, the results indicate that VOT has high influence in pointing out the differences among speakers specifically in speech tempo, so they normalized the tempo factor and found speaker differences even after normalization. Moreover, Smiljanic & Bradlow (2008) who investigate the effect of different speaking rates in different speaking styles -clear speech and conversational speech- on vowel and stop consonants in English and Croatian languages. The results indicate

https://phonetics-acoustics.blogspot.com

that VOT duration was longer in clear slow speech than in conversational speech in both languages, which reveals decreasing speaking rate. Also, Huggins (1972) in a research about the perception of temporal phenomena in speech; declares that vowel-onset hypothesis explain almost all of results of the experiment, and he also suggests that some vowels onsets are more important than others, adding that importance increases when the vowel is stressed or when it is the first vowel after a word boundary.

On the other hand; Arvaniti (1999) studies the effects of speaking rate changes on the timing of single and geminate sonorants. Sonorants display the relatively loose timing of vowels, while the contrast between geminate and singletons relies mainly on duration. Arvaniti's experiment includes a comparison of single and geminate sonorants /m, n, l, r / in Cypriot Greek and Standard Greek. Eight speakers are asked to read sentences from cards which contain the sonorants in four disyllabic minimal pairs in intervocalic positions. Each speaker produced six repetitions of the test sentences at normal and fast rates. The results indicate that the duration of all sonorants (single and geminate) is shorter at fast rate than at normal rate, except /m / & / r / was not affected by rate changes. Those results coincide with Salem (2006) who studies the effect of speaking rate changes on the production of singleton and geminate consonants in Colloquial Egyptian Arabic. Her experiment includes 26 minimal pairs where the target consonant was in an intervocalic position. Four native speakers are asked to read the target sentences at normal and fast rates. The results conclude that all geminate consonants are longer in duration than singleton in normal and fast speaking rates, where the duration of singleton and geminate consonants decreased in fast speaking rate than that in normal speech. Moreover, increasing speaking rate do not influence the difference in duration between geminate and singleton consonants.

The speaking tempo is influenced also by ***the syllable structure*** of an utterance. Whereas, syllable structure is influenced primarily by the whole syllable duration, and respectively, syllable duration is positively related to syllable complexity[11]. The more complex the syllable is, the greater the syllable duration; and conversely, the speaking tempo will be slower (Miller et al. 1984).

Miller& Baer (1983)[12], examines how syllable duration and transition duration of the initial consonants in the syllables / ba / and / wa / influence the rate of speaking. The speakers were asked to produce these

[11]Complex by meaning of has more segments.

[12]Cited in Miller, M., Grosjean, F. & Lomanto, C. (1984) Articulation rate and its variability in spontaneous speech: reanalysis and some implications. Phonetica 41, pp. 215- 225.

https://phonetics-acoustics.blogspot.com

syllables at different rates ranged from very fast to very slow. The results indicate that the speakers can produce syllables ranging from under 100 msec. to over 600 msec. And on the perceptual phase, transition duration is considered one of the most important cues that listeners depend upon, i.e., as the duration of the syllable increases, increasingly longer transition duration is required to perceive it. Also, Dankovicova (1999) studies the articulation rate variation in Czech and English, and discusses that the articulation rate of words was also significantly affected by their size in syllables and their word class because of the tendency of the word articulation rate to increase with increasing number of syllables.

Conversely, the alternations in the rate of speaking also modify the structure of syllables, so as to affect many of segmental properties for consonants and vowels (Miller et al. 1984). Therefore, the role of speaking tempo in ordering segments is most obviously important in describing acoustically complex syllables, because certain segmental properties are associated with their position within the syllable (Lisker 1989).

Another linguistic factor that affects the global speaking tempo is ***utterance length***. The relation between duration and length is felt to be so much closer. Utterance length is the simplest factor and its direct influence is intuitively clear; the longer the acoustic segment is, the slower it sounds and vice versa. Also, the stretch of speech must be taken into consideration. Because, the first part of an utterance can be spoken fast, while the second part can be rather slow, or vice versa. Accordingly, an average rate calculated for an utterance does not necessarily reflect the tempo characteristics of different parts (Trouvain 2003).

Kohler (1986) investigates parameters of speech rate perception in German words and sentences according to the overall duration. Short and long meaningful utterances were played to a group of seven native German listeners to classify the rate of speaking (slow or fast) of those utterances. The results of the experiment show that the speech rate perception depends on the overall duration. Therefore, long utterances sound slower than the corresponding short ones that contain the same number of syllables. Kohler also explains that, perceptually, the perception of tempo reflects what happens in production; when utterances are slowed down or speeded up, their component parts are lengthened or shortened, in accordance with slower or faster movements of the articulators, thus increasing or decreasing the overall duration.

On contrary, Slis et al. (1991) studies the development of speaking rate in Dutch children in relation with age and utterance length in spontaneous speech. The experiments include 40 Dutch children divided into four groups according to their age. The analysis of the data clarified that the rate of speech is higher in the long utterances than in the short

https://phonetics-acoustics.blogspot.com

ones in all age groups. Also Schwaba & Avanzi (2015) examines the articulation rate in read speech and in spontaneous speech in seven variants of French. The results indicate that the longer the utterance is, the faster the articulation rate becomes.

Co-articulation is one of non-linguistic ways in which speech samples can differ with respect to the rate of speaking. Co-articulation occurs during connected speech, whereas, the movements of different articulators overlap in time and interact with one another for the production of successive phonetic segment because the vocal tract configuration at any point in time is influenced by more than one segment. Consequently, the degree of co-articulation can vary as a function of speech rate. Along with co-articulation, assimilation processes occur where neighboring sound segments that require the same articulator use a single articulatory gesture for both sounds (Farnetani 1990: P. 371).Moreover, Beckman (1989) states that co-articulation occurs in connected speech in casual or fast speech and in fluent read speech as well. It can occur within the word, as well as, at word boundaries. Which can cause a drastic restructuring of the spectral characteristics of segments at the edges of words. Final segments can change by assimilation according to the initial segment of the following word, and they can even be seemingly deleted.

Slis et al. (1991) assumes that the speaking rate is affected by the degree of co-articulation. This assumption is tested on spontaneous speech of Dutch children correlated with their age. The results conclude that high degree of co-articulation leads to fast speaking rate; where in unstressed vowels are generally reduced and may even be completely deleted.

Recasens (1985) analyses acoustically the co-articulatory effects of V-to-C and C-to-V in CV sequences in Catalan for 18 consonants and 8 vowels. The results indicate that Catalan phonemes differ in the degree of resistance to co-articulation due to the differences in the degree of articulatory constraint which are involved in the production of gestures for adjacent phonemes.

Also, Bona et al. (2008) studies the co-articulation rules in different speaking styles (reading, interpreted, and spontaneous speech) in Hungarian. It is investigated whether the co-articulation is affected by speaking style, articulation rate, or reading fluency. The results show that the articulation rate and the reading fluency do not affect co-articulation at all, whereas speaker variability and different speaking styles have small effect on co-articulation. Therefore, they suggest that co-articulation is not only phonologic – phonetic research but also a psycholinguistic and sociolinguistic study to consider the individual speaker variability.

https://phonetics-acoustics.blogspot.com

2.3.3. Tempo and Supra-Segmental Features

Connected speech is characterized by continual modifications or alterations in ***F_0 movements (***which is realized perceptually by ***pitch & intonation), stress (***which is realized perceptually by ***loudness & intensity),*** and ***pauses*** (which are called supra-segmental features, and they are necessary for speech communication). Any changes or alternations in these supra-segmental features directly influence the rate of speaking and its perception in some way.

Li & Zu (2008) studies the effect of speech rate on discourse prosody in Standard Chinese. A female non-professional speaker was asked to read a Chinese narrative discourse (of four paragraphs) with three different speaking rates (normal, fast and slow). The results of the experiment indicate that the increasing and decreasing of speaking rate will reconstruct the prosodic structure of the discourse in several ways as the following; 1) the number of prosodic boundaries decreases by 0.8 times at slow rate, and increases by 1.3 times at fast rates. 2) Speaking rate variation affects the overall F0 value. Increasing speaking rate will raise F0 value, and decreasing speaking rate will lower F0 value. 3) The silent pauses decrease, as the speaking rate increases. When rate increases, the silent pause is shortened by about 0.8 times; and lengthened by 2.6 times when rate decreases. 4) Different speaking rates have a significant influence on stress distribution on Chinese discourse. For fast rate, the stressed finals and syllables are shortened, whereas slow rate has no effect on stressed finals and syllables except lengthening the stressed syllable.

With regard to ***pitch***; Kohler (1983)[13] states that "F_0 increases with increasing speech rate." On the contrary, Igarashi (2004) tests the effect of F_0 movements (pitch raising and falling) on speech rate in Russian. Russian sentences were read aloud at three different speaking rates (normal, slow, fast). The results show that F_0 increases with slow speaking rate in Russian languages. Those results coincide with Trouvain & Grice (1999) which studies the effect of different speaking tempi on the prosodic structure of German. The experiment includes three female native speakers; each produces three readings of German texts at three different rates (normal, fast and slow). The results indicate that increasing tempo leads to decreasing F_0 and pitch range and vice versa.

Whereas, Casper & van Heuven (1991) investigates the aspects of pitch movements in speech rate in Dutch. The experiment includes six Dutch speakers who read aloud texts at normal and fast speaking rates.

[13]Cited in Kohler, K. J. (1986) Parameters of speech rate perception in German words and sentences: duration, F_0 movement and F_0 level. Institut für Phonetik, University of Kiel. Language & Speech, 29: 115–139.

https://phonetics-acoustics.blogspot.com

Two phonetically trained listeners marked the pitch positions and boundary movements of the materials. The results showed that 40 % of pitch movements are disappeared at the fast speaking rate.

Also, Fougeron& Sun-Ah (1998) investigates the speech rate effects on French intonation according to the phonetic realization of F_0 contour and the prosodic organization of a text. The experiment includes three French speakers who are asked to read a text at normal and fast speaking rates. The results show that fast speaking rate includes reduction of pitch ranges and movements, and also the prosodic organization of intonation at phrase boundaries are reduced or deleted.

So, pitch has already been found in the production of experiments to be correlated with speaking tempo. And also, *perceptually*, the perception experiments have shown that the listeners judging high pitch production as fast speaking tempo and vice versa.

Kohler (1986) studies the perceptual effect of F_0 movements on speech rate in German. The experiment includes German tokens produced with identical duration but with high and low F_0 in the final position. Eight native German listeners were involved to determine whether those tokens were spoken in slow or fast speaking rate. The results indicate that, on the perceptual level, higher F_0 sounds faster while lower F_0 sounds slower.

In respect of ***stress***, a stressed syllable in a word is generally spoken with more articulatory force, resulting in a syllable that is louder, longer in duration, and higher in pitch than an unstressed syllable. This is proved experimentally by Zhang et al. (2008) which studies the acoustic characteristics of lexical stress in English. The study includes a comparison between native English and native Mandarin speakers about the use of stress in correlation with F_0, duration and intensity. The results of the investigation conclude that the acoustic analysis of both native English and native Mandarin speakers shows that they produced the stressed syllable with higher F_0, longer duration and greater intensity than unstressed syllable. Also, Van Son& Pols (1992), studies the vowels formant movements in Dutch at fast and normal speaking rates. The results show that the stressed tokens were 30 % longer in duration than the unstressed ones for both rates.

Concerning ***pauses durations and their frequency of occurrence***, the main thing that contributes to the fluency of speech is the number and duration of pauses a speaker makes. It is clear that people pause when they speak, and that they do this for several reasons. A pause may refer to a rest, hesitation, or temporary stop. It is an interval of silence and may vary in length. The speaker may use pauses to enhance the message delivery, to emphasize that the information coming next is important, or

https://phonetics-acoustics.blogspot.com

to give the audience time to process what he/she have just said (Rose 2002).

According to Laver (1994) and also Zellner (1994)pauses can be divided into ***filled pauses*** which contain some pronounced non-linguistic material such as *er, um, mm* and ***silent pauses*** which contain a silent expiration or can be seen as a silent portion in the speech signal.

In addition, Rose (2002) defined another type of pauses which are called ***Hesitation pause.*** Hesitation pause is one of the linguistic pauses and can refer to a syntactic pause. Hesitation pause is syntactically irregular and occurs within a clause to indicate some cognitive (verbal) planning in process or for the speaker's brain in order to make decisions in the episodes of speech, and might not occur at an intonation phrase or a syntactic phrase boundary. Hesitation pauses also are considered as a very important, and only type of psychological pause which can be *silent, filled, use prolongation, repetitions, false starts, retracting* or *manifest combinations of these*; therefore, such pauses reflect the emotional states of the speaker like tension and anxiety (more tension involves more occurrences of hesitation pauses). Hesitation pauses occur in spontaneous speech and reading as well. Hesitation pauses also imply different styles of individuals speaking, different social interactions and different characterizations of individuals' personalities.

Künzel (1987)[14] emphasizes the importance of hesitation pauses and its repetitions particularly in forensic speaker identification FSI, and stated that; "In fact the phonetic realization of hesitation phenomena is highly speaker specific, not only in its phonetic quality, but also under certain circumstances with respect to distribution, i.e. it frequency of occurrence particularly in relationship to the number of so-called silent pauses ..."

Künzel (1997)[15] quantifies *the degree of hesitancy of a speaker* shows as *the proportion of filled pauses to all pauses*, under the assumption that the proportion of filled to silent hesitancy pauses remains constant and the resulting proportions tended to show large differences *between* speakers and relatively small *within*-speaker differences and are therefore possible forensic-phonetic parameters.

To vary the tempo, speakers change the number of pauses as well as the duration of pauses. In fast speech, many of the pauses which might be observable at normal speed are temporally reduced or completely omitted, which indicates that there is a significant correlation between speaking tempo and pauses. Trouvain & Grice (1999) while

[14]Cited in Rose, P. (2002) *Forensic Speaker Identification.* London, UK: Taylor and Francis.

[15]Cited in Rose, P. (2002) *Forensic Speaker Identification.* London, UK: Taylor and Francis.

https://phonetics-acoustics.blogspot.com

studying the effect of different speaking rates on the prosodic structure in German, found that all speakers increase the number of pauses (adding) and their durations at the slow rate of speaking; and reversely, reduce the number of pauses (deleting) and their durations at the fast speaking rate. Also Lane & Grosjean (1973) examines the perception of five different reading rates by speakers and listeners to investigate the extent to which the speaker changes the articulation rate and the frequency of pauses and their durations; in order to produce an overall change in reading rate. The results showed that when a speaker varies his rate of reading, he primarily adds or subtracts pauses as the same duration as at the strategic syntactic locations, and alters the articulation rate and the pauses durations much less.

On the other hand, Fletcher (1985) studies numbers of factors that influence the speaking rate in connected speech. The experiment includes native French speakers who are asked to read a passage at increasing tempo (from slow, to normal, to fast). The analyses of her results showed that speakers modify their pause behavior substantially. All speakers reduced their percentage of pause to total speaking time from slow to fast speaking tempo. Also the number of pauses inserted in each reading decreased from the slow to the fast versions, and also pause length decreased from slow to fast rates. Moreover; Fletcher concluded that pause duration plays a much greater role than articulation rate AR in determining the changes in the overall speaking tempo.

Whereas; on *the perceptual* level, Grosjean & Lane (1974) investigates experimentally the effect of the number of pauses and their durations and the articulation rate in altering the rate of speaking in reading. Their experiment includes 17 graduated students who are asked to listen to a read passage of 51 words to measure the perception of speaking rate in a scale. The analysis of the results indicates those pause frequency and articulation rates are equally important in their effect on changing the speaking rate and its perception as well. Sequentially; according to listeners' judgments, the speech sample which is characterized by more pauses and longer pauses is judged to be spoken in slow speaking tempo compared to speech at normal tempo, and reversely, the speech sample which is featured by fewer pauses and shorter pauses is judged to be spoken in fast speaking tempo compared to normal (Trouvain2003).

Pauses are one of the most important cues that have a strong effect and function in chunking information in speech because the prosodic phrase boundaries are very often marked by an acoustic pause. Fors (2011) hypothesizes that pauses that occur between syntactic units should be longer than pauses that occur within syntactic units because pauses between syntactic units include turn-talking negotiation. Fors's

https://phonetics-acoustics.blogspot.com

experiment includes five female speakers, speaking spontaneously in pairs, recorded for 10 minutes, and all pauses are detected and calculated for each speaker. The results of Fors's experiment indicate that three of the five speakers are in line with her hypothesis, whereas one speaker shows no difference and one speaker made longer pauses within syntactic units.

Last but not least, tempo variation in speech production is dependent on many variables which are not only phonetic or phonological in nature, but also include other factors like emotions, attitudes, stress, age, speech and hearing impairments, the communication partner, and habitual speech rate, etc. Thus, a change of speaking tempo results in changes at many levels, all of which can be more or less closely linked to each other. Of course, there is still much to discover about the domains in which these factors operate.

https://phonetics-acoustics.blogspot.com

2.4. Listener's perceptual abilities of speaking tempo.

An acoustic analysis on its own is inadequate because there is a close relationship between the inherent properties of the speech output mechanism and the perceptual recognition routine that makes the perceptual abilities of humans in identifying speakers' voices an important area that can't be easily ignored.

In general, humans are frequently able to identify familiar speakers without seeing those (Bonastre et al. 2003). For instance, when they are speaking outside the door before knocking (Nolan 2001), or while answering the telephone and recognizing the voice of the caller, perhaps after only one or two syllables, in spite of the degradation of the acoustic signal imposed by the limited bandwidth (300 - 3,500 Hz) of the telephone line (Watt 2009). Naive listeners can also estimate some kind of personal information of unknown speakers (like gender, age, weight and height) about the speaker with remarkable accuracy; depending only on the speaker's voice sample; moreover, people have recognized voices they have not heard for substantial periods of time, for example, like identifying the voice of a presenter of a television programmer forgotten since childhood, or that of a schoolmate with whom contact had been lost for decades. Even if we cannot immediately name the person we think is talking, we may feel absolutely certain that the voice is known to us.

Yarmey (2001) describes experimentally ear witness accuracy in speaker identification. The experiment was designed to examine the accuracy of the perceptual abilities of listeners based on aural / auditory speaker identification through using their ears. The results reveal the fact that listeners can identify speakers perceptually with remarkable accuracy and better than chance. However, there are no significant differences in speakers' performance. Hence, those results indicate that there are adequate perceptual cues in the voice of the speaker which reflect, to some extent, information about that speaker and revealing his/her identity. Because of that, those extraordinary perceptual abilities of the listeners in identifying speakers must be considered and utilized in forensic phonetic field, more precisely for forensic speaker identification FSI.

According to Laver (1994), there are only four perceptual domains / cues (which correlated to each other) available to the human auditory system for differentiating the elements of speech: ***1) perceptual quality, 2) duration (tempo), 3) pitch*** and ***4) loudness.***

Along with ***the perceptual quality, pitch*** and ***loudness***, many researchers proved experimentally that listeners have a distinct ability in identifying speakers' voices, even under the most difficult circumstances of similarities (in anatomical structures, speaking styles or imitating

https://phonetics-acoustics.blogspot.com

someone's speaking style). This makes Eriksson (2005) inquire whether the poorer sound quality of recorded telephone conversations adversely affects the speaker identification process by ear witness, and if so, to what extent and how. There are surprisingly few studies that address this question, even though, there are some results which indicate that the problem might not be as serious as one might expect. Also, Hei-Yin (2004) studies the forensic speaker identification under optimum noisy environment and over telephone condition; for the acoustic analysis, where the formants were masked and removed by background noise and telephone system, so spectrographic patterns were greatly distorted. Thus the paired-spectrographic comparison could not be made. The lack of sufficient information made acoustic speaker identification impossible under unfavorable conditions. Reversely, on the perceptual identification, the results indicate that the accuracy of perceptual identification was around 80% under optimum condition, implying that masked or removed high frequency formants, that is, F2 and F3, which carry information of vowel distinction, are not the only cues for people to recognize speaker’s voice. But also, pitch, speaking rate and intonation, which would not be masked or removed by noise and telephone system, are considered as hints for ear witnesses as well. Thus, the ear witnesses’ performance is not affected under different conditions. Furthermore, Foulkes & Clark (2007) describes an experiment about the effects of electronic vocal disguise on speaker identification. They found that the results show significant effects for both listener and speaker. Furthermore, Johnson & Azara (2000) investigates the perception of personal identity of twins' speech. The experiment is designed to discover whether listeners could detect differences between twins' speech because identical twins have virtually identical vocal tracts and grew up together in the same home. The results indicated that listeners' sensitivity of twin's speech differences is greater than chance, and stable over changes in experimental conditions. All of the aforementioned studies and their results have a significant importance because it is documented that listeners can detect the speakers' differences which are not due to differences in the vocal tract or formative language experiences; and also because they are in consistence with the hypothesis that there are individual differences in speech production due to idiosyncratic habits of pronunciation in addition to differences due to vocal tract anatomy and also supra- segmental features of the speaker's voice.

Lindh (2009) supposes that listeners use at least two different strategies to identify speakers perceptually. They may pay a lot of attention to *voice quality* or concentrate on *speaking parameters* such as *articulation rate* and maybe to some extent pronunciation.

In respect of ***tempo (duration)***; identifying a speaker perceptually is a difficult task, so speaking tempo is an extremely useful parameter for

https://phonetics-acoustics.blogspot.com

listeners' perception when comparing speakers, especially for forensic applications. Laver (1994: P. 431 & 432) pointed out that listener's perception is capable of discriminating different durations. The human auditory system is psychophysically capable of registering very fine temporal differences of duration under favorable experimental conditions. The duration of individual speech-segments ranges from about 30 msec. to some 300 msec., whereas the psychophysical threshold for just-noticeable differences in duration between two such sounds is of the order of 10 to 40 msec.

Furthermore, it is traditionally common that as much as the rate of speech is increased (beyond the normal limits), the articulation is deteriorated and also the intelligibility (Vingerling 1984). Organs of speech cannot produce the speech as fast as it is processed in the brain, and consequently this results in losing of speech information (some speech sounds) and sequentially may have a bad effect on the listeners' perception. However, Laver (1994) states that the capacities of perception exceed those of production and also the maximum rates of perception are faster than the maximum rates of production. Lane & Grosjean (1973) examine the perception of reading rate by speakers and listeners. Their experiment designed to examine the speaker's perception of his own rate of reading; through 12 graduate students who read an experimental passage of 51 words at different five rates. The results showed that when a speaker doubles his reading rate he perceives a sixfold increase, whereas a listener perceives less than threefold increase. Lass, Mertz & Kimmel (1978) investigates the effect of temporal features on speaker sex identification. The experiments include 20 speakers who recorded four sentences. The data are constructed from these recordings to represent three experiments (forward-played, backward-played and time compressed). Thirty naïve listeners participated for each experiment and asked about the sex of the speaker. The results indicate that listeners' judgments were better than expected for each three experiments.

Also, Bonastre et al. (2003) states that the perceptual abilities of the listeners, especially their perceptual sensation of different tempi, may depend on some factors, which influence their perceptual ability and could be responsible for that remarkable accuracy in, identify both different speakers and different speaking tempi as well. Some of those factors can increase and strengthen this perceptual ability, such as the familiarity with the speaker, the duration of the sample or the context, the lack of vocal stress and disguise, and also training.

The first factor that influences the perceptual sensation of speaking tempi is the ***listener's age***; Watt (2009) assumes that these great perceptual abilities of listeners in identifying speakers seem related to the listener's age. This is also proved by Smith (1978) who investigates that

https://phonetics-acoustics.blogspot.com

the rudiments of timing systems are present even in 3 years old children. Also Miller (1981) discovers by evidence the presence of a perceptual awareness of pre-articulate infants; they seem to be able to process transitions and syllable durations in different speaking rates in categorical manner similar to that of adults. While Mann et al. (1979),reports that only those children in their sample over 10 years of age could identify speakers at adult-like levels of accuracy (these findings have clear implications in forensically relevant scenarios involving child witnesses, whose testimony must in any case be treated with particular caution (Parker et al. 2006)). Moreover, Clifford et al. (1981) finds that listeners between the ages of 16 and 40 perform better in speaker identification tasks than older (40+) listeners.

The second factor is ***the quality of the voice sample***. Hei-Yin (2004) points out that, while identifying unknown speakers' voices, background noise may affect the listeners' perceptual abilities as well as their confidence about their choice. Whereas, the results conclude that the conditions of noisy environment and telephone do not have negative effect on the perceptual abilities of the listeners during speaker identification process. Sequentially, these remarkable results indicate that there is supra-segmental information such as pitch, speaking rate and intonation also play a part in identifying speaker’s voice. However, listeners’ confidence in speaker identification is not a reliable predictor of accuracy of identification and their confidence is greatly affected by psychological factors rather than their abilities. Moreover, acoustic analysis is possible only under favorable conditions.

The third factor is ***the listener's native language***, regarding unknown speaker identification, listeners' performance is generally worse when they do not speak the same language as the target speaker and it remains unclear whether this effect is dependent on the relatedness between listeners' native language. Schiller & Koster (1996) discusses this experimentally in a report about evaluation of foreign speaker in forensic phonetics, where it is found that listeners perform generally better in identifying unknown speaker when they have a knowledge of the spoken language, adding that, speaker identification does not only involve purely phonetic features, but also incorporates linguistic information. Also Perrachione et al. (2011), in its research about the human voice recognition depending on language ability, state that the perceptual abilities of human listeners are more accurate in identifying speakers' voices when they can understand the language being spoken. It can be stated that the native language background of the listeners may play a role influencing their perceptual abilities in identifying unknown speakers. Moreover, in respect of identifying different speaking tempi, Vaane (1982) assumes that the listeners' perception of different speech tempi of different languages is influenced mainly by their lexical knowledge about

https://phonetics-acoustics.blogspot.com

that language. Yet, Vaane (1982) supposes that the perception of the speech tempi of the unfamiliar languages is faster than the perception of different speech tempi of the listener's mother tongue; i.e., when the listeners estimated different speech rates, they were influenced by their lexical knowledge of the language. The results, unfortunately, came to weaken his supposition; indicating that the listeners perceive their main cues for tempo detection depending on some temporal features in the speech signal, and they did not relay on their knowledge of the lexical information of the language.

The forth factor includes ***the hearing impairments listeners***. Experimentally, there is no significant difference between normally hearing and hearing impaired listeners in perceiving different speaking tempi (Vingerling 1984). Vingerling's results conclude that normally hearing and hearing impaired listeners can recognize the normal and the slow speech rates with high scores; in contrast with fast speech rate, which makes Vingerling wonders about two important points: 1) what kind of factors play a role in the realization of speech rate?, and, 2) are those factors as same as those that are involved in the perception of speech rate and its influence on intelligibility? (See chapter 4 section 4.3. you may find suitable answers).

Jessen (2008) adopts ***the holistic approach***. Intrinsically, listeners have a certain ability to recognize different voices, rather than different speakers. They are not trained in the systematic analysis of language and speech; their perceptional ability is essentially holistic. Holistic, literally, means considering the whole thing, whereas technically for the auditory-perceptual domain, it means to listen to speech without trying to analyze its components, while focusing mentally on the question whether two speech recordings have been produced by the same speaker or by different speakers. Jessen's suggestion relies on Gestalt theory, "*the whole is greater than the sum of its parts*", and he would apply this theory to speaker identification. In fact, it has been shown that the speaker identification performance of holistic methods is clearly good enough to be of practical use. Alexander et al. (2005) compares experimentally aural and automatic speaker recognition in the context of forensic analyses. The experiments include 90 phonetically untrained subjects simulated by several forensic cases. They estimate the strength of evidence for both humans and the automatic system. The results show that the performance of humans are highly significant and they analyze the perceptual cues that human listeners rely upon in identifying speakers in order to improve the accuracy of the automatic systems. Also Li & Zu (2008) while studying the effect of speech rate on discourse prosody in Standard Chinese, 10 untrained subjects were asked to identify the three speaking rates for the recorded discourse. The perceptual results indicate that all the involved listeners got 100 % correct rates.

https://phonetics-acoustics.blogspot.com

Jessen (2008) suggests that this holistic approach can be applied more quickly than analytical speaker identification and more suitable for naïve listeners in identifying speakers; so as most of naïve listeners followed this approach in order to do the process of speaker identification. Some studies have shown that phoneticians are somewhat better in holistic speaker identification than naïve listeners. However, holistic methods do not neither fully satisfy the scientific interests of the phonetic expert nor the demands of the court.

As clarified, as well as speaker's voice conveys considerable information about the speaker, listeners, sequentially, perceive and realize this information in order to utilize it in evaluations and attributions.

https://phonetics-acoustics.blogspot.com

مكتبة وملتقى علم الأصوات
phonetics-acoustics.blogspot.com

CHAPTER 3

3.1. Procedures

3.1.1. Data collection

The present experiment includes 10 unknown speakers (5 females and 5 males) of colloquial Arabic language, with no recorded speech disorders. Speaker's ages estimated between 19 to 40 years old. Natural spontaneous speaking style is elicited for 30 seconds for each speaker trying to avoid the effect of any stress or the domination of any specific emotion.

All the data are collected through a radio program called "The Press in Their Eyes الصحافة فى عيونهم" which is a daily program in Radio of Alexandria (Bakous Alex, frequency 101.1). The host of the program goes down to street everyday and interviews a street person. This person could be a male or a female who is reading any daily newspaper and whose identity is unknown to the host and to the listeners. The host asks a simple question; which is: *what's your comment on one of the news that you have read at that daily journal?* Then, the unknown speaker starts to talk spontaneously without any recommended preparation about any topic that he or she chooses. Accordingly, that unknown speaker is an Alexandrian citizen who may have received intermediate education or may be well educated.

3.1.2. Recordings

The data are collected and elicited by the program host who asks the unknown speaker about his/ her comments or opinions about any piece of news of the daily journal headlines. The whole duration of each episode is about 5 minutes for every speaker; directly recorded from the radio channel using ***Samsung mobile phone recorder as wav. files***; to avoid any transmission distortions.

Then, all the episodes (10 episodes of 10 unknown speakers, each of which is 5 minutes) are transmitted into a laptop device for editing.

Therefore, the researcher used cutter software for cutting only 30 seconds of continuous and spontaneous speech of each speaker from the whole speaking time (from the whole episode which is 5 minutes).This cutter software is called ***"Easy audio ogg wma wav cutter software***" and downloaded from the website: www.koyotesoft.com

At last all the edited data (only 30 seconds of spontaneous speech for 10 unknown speakers) are exposed to ***Praat software*** (www.praat.org) for the analysis (next step).

https://phonetics-acoustics.blogspot.com

3.1.3. Analyses

All the data are analyzed manually with the aid of Praat softwarefor all speakers. The analysis procedure is composed of three sequential steps.

The first step is the transcription process in which every 30 seconds of recording spontaneous speech for each unknown speaker are phonetically transcribed by using IPA symbols. The researcher transcribed all the data manually through careful listening depending on the ears of the researcher with the aid of Praat software as a listening tool. Broad transcription type is used for this research because the main concern of this transcription process is counting the number of the pronounced syllables in a particular time (which is 30 seconds of spontaneous speech for each informant). So, it will not be of great importance how an informant pronounces a particular phoneme as long as his / her pronunciation does not affect the number of the pronounced syllables (see appendices).

The second step is the segmentation process which means dividing the transcribed speech into syllables; this process is done manually by the researcher (see appendices).

The third step is the calculation process in which speech rate SR and the articulation rate AR are calculated with their durations. Also the number of pauses and the duration of each pause are counted too (see appendices).

3.1.4. Measurements

Listed below are all the acoustical measurements illustrated with their mean of calculation for all the ten unknown speakers:

- ***Fundamental Frequency F_0***is measured for all speakers using Praat voice report.

- ***Intensity*** is measured for all the speakers with Praat software through getting the mean intensity.

- ***The number of the pronounced syllables*** for each speaker; how many numbers of syllables the speaker has pronounced in only 30 seconds. The number of the pronounced syllables calculated manually by the researcher through counting all the produced syllables after segmentation process.

- ***Speech Rate SR*** is measured according to the following definition "the number of syllables per second including the

https://phonetics-acoustics.blogspot.com

whole speaking time (with all pauses and hesitations)" which is 30 seconds for each speaker.

- ***Articulation Rate AR*** is measured according to the following definition "the number of syllables per second excluding the pause time and all the hesitation duration". Note that the excluded pause time and hesitation duration will vary from one speaker to another.

- ***All Pauses durations*** are measured by combining the duration of each pause in each speaker's utterance with the duration of each pause between utterances.

- ***The number of pauses*** for each speaker is counted manually by the researcher through counting the number of all pauses (filled and silent) which occurred in the whole speaking utterance (30 seconds for each speaker).

- ***The duration of each pause*** that occurred in the whole speech sample (in 30 seconds) for each speaker with the aid of Praat software, determining the type of each pause.

- ***Percentage of pause time*** is measured manually by the researcher, through calculating the proportion of all the pauses time (the duration of all pauses) to the whole time of the speech sample (which is 30 seconds).

- ***The degree of hesitancy*** is measured manually by the researcher for each speaker through calculating the proportion of filled pauses to all pauses for the overall speech sample.(See appendices)

https://phonetics-acoustics.blogspot.com

3.2. Perceptual Test

Sixty listeners of university students aged between 17 and 25 years old, with no recorded history of hearing impairments.

Each listener was sitting directly in front of a laptop computer device with approximately three feet distance. The listeners were listening to the voice line-up (mp3 playlist, with 2 seconds interval between each informant and the following) through a loud speaker (attached to the laptop computer device) which was set up on medium volume.

The listeners received some instructions from the researcher for doing the perceptual test perfectly:

1. Each listener received a listening sheet (see Figure 2) which contained the ten unknown speakers (5 females and 5 males listed one by one) titled as informant 1, informant 2, …………., informant 10.
2. The listeners were asked to listen carefully to the voice line-up of the ten unknown informants three times, at most, in order to enable them to select the fastest speaker and the slowest speaker.
3. Then, each listener selected the fastest speaker and the slowest one by marking (√) in front of his or her title at the listening sheet (see Figure 2).

Observe that, the ten informants' voices were listed one by one, male followed by female, in the voice-line up in order to distract the listener's attention from the gender of the speaker, since almost all acoustic measurements and perceptual expert descriptions show experimentally that there are no significant differences in speech tempo between men and women. In other words, tempo of speech has no relation to the gender of the speaker (see section 2.3.1 tempo and physical characteristics).

Speakers المتكلمون	The FASTEST الأسرع	The SLOWEST الأبطأ
Informant 1		
Informant 2		
Informant 3		
Informant 4		
Informant 5		
Informant 6		
Informant 7		
Informant 8		
Informant 9		
Informant 10		

Figure 2: the listening sheet; where the involved listeners are marking (√) in front the fastest informant and the slowest one too.

https://phonetics-acoustics.blogspot.com

https://phonetics-acoustics.blogspot.com

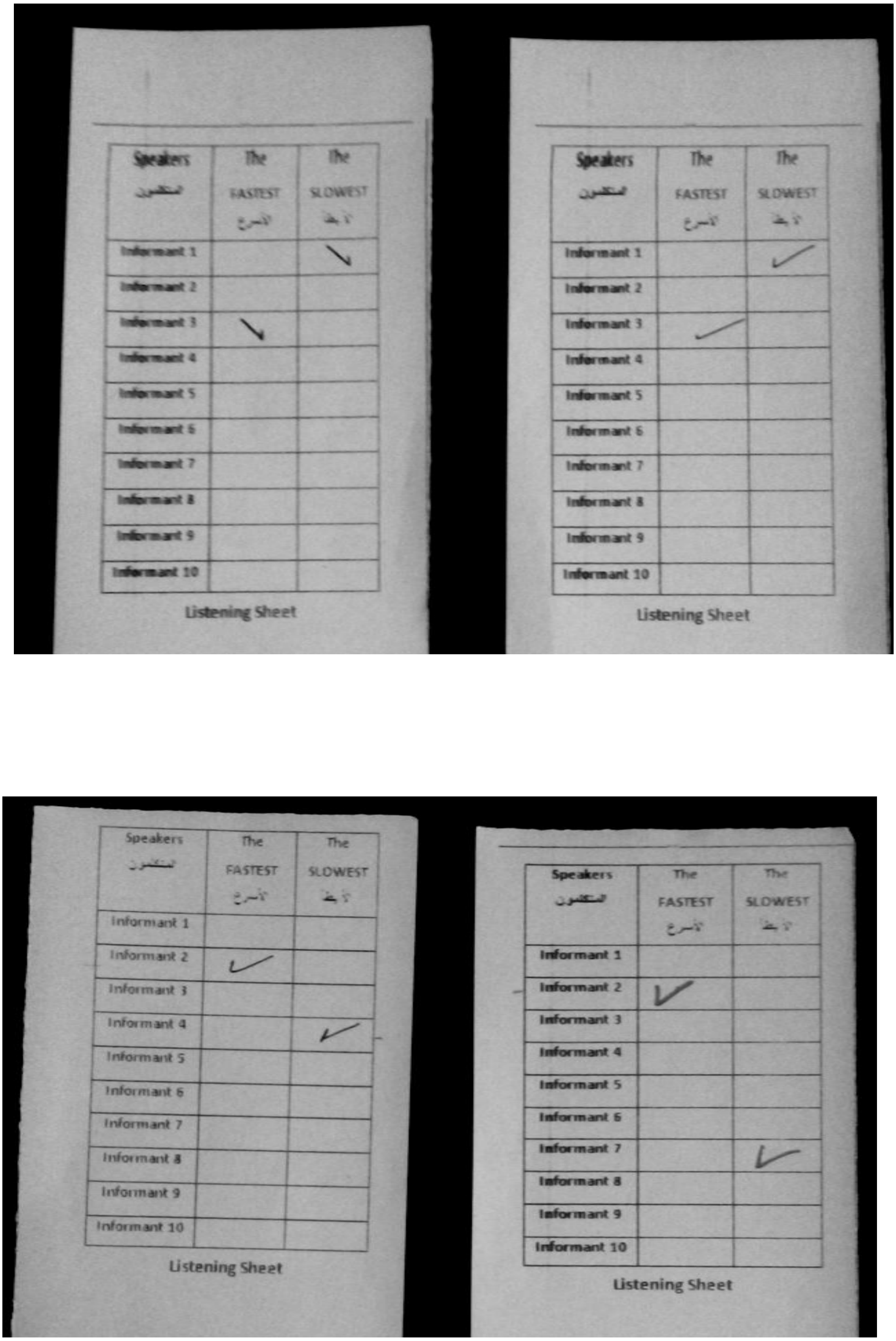


Figure 3:Samples of the listening sheet after the selections of the listeners who involved in the perceptual test.

CHAPTER 4

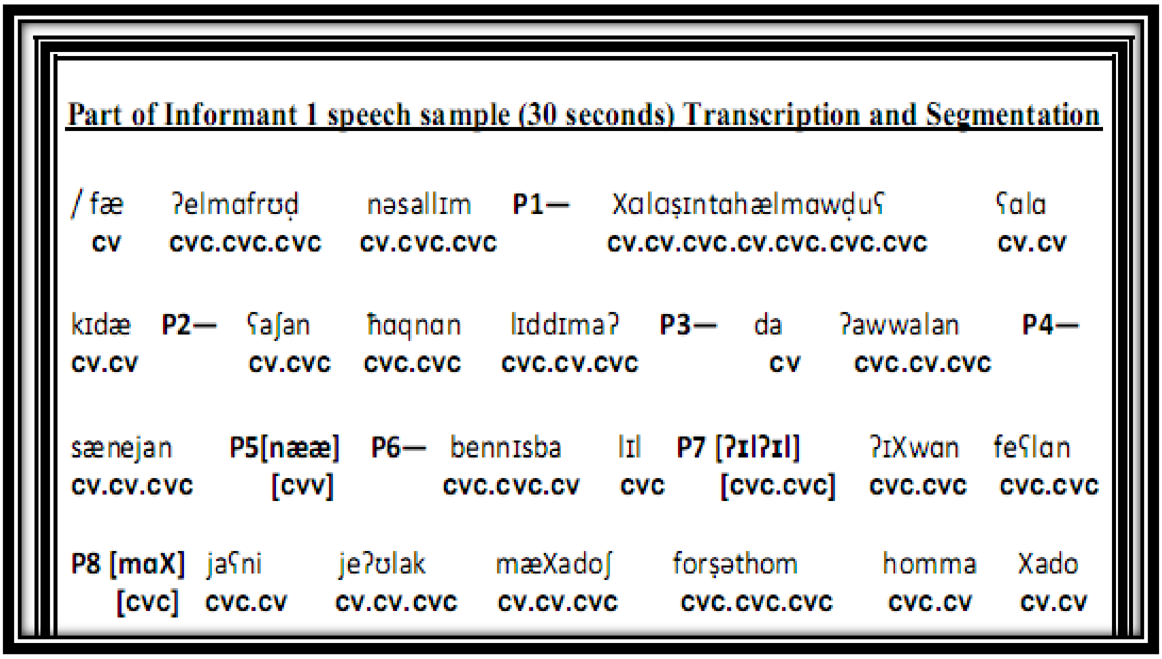
4. Results and Discussion

This experiment is designed in order to be utilized in forensic case work to identify unknown speakers depending on their tempo of speaking (rate of speaking) on the basis of two approaches: 1) an acoustic approach; based on acoustical measurements and statistical analyses; and 2) a perceptual approach; based on the ability of the naïve listeners to recognize the differences in speaking tempi; specifically identifying *the fastest* speaking tempo and *the slowest* speaking tempo depending on their ears.

4.1. Results of analyses and acoustical measurements

a) Transcription and segmentation processes

Figure 4 is an example of a manual broad transcription of part of speaker 1's spontaneous speech[16]. It includes also a segmentation process i.e.; separation between the uttered syllables (or segregation).


Part of Informant 1 speech sample (30 seconds) Transcription and Segmentation

/ fæ ʔelmafrʊḍ nəsallɪm **P1—** Xɑlɑṣɪntahælmawḍuʕ ʕala
cv cvc.cvc.cvc cv.cvc.cvc cv.cv.cvc.cv.cvc.cvc.cvc cv.cv

kɪdæ **P2—** ʕaʃan ħaqnan lɪddɪmaʔ **P3—** da ʔawwalan **P4—**
cv.cv cv.cvc cvc.cvc cvc.cv.cvc cv cvc.cv.cvc

sænejan **P5[nææ]** **P6—** bennɪsba lɪl **P7 [ʔɪlʔɪl]** ʔɪXwan feʕlan
cv.cv.cvc [cvv] cvc.cvc.cv cvc [cvc.cvc] cvc.cvc cvc.cvc

P8 [mɑX] jaʕni jeʔʊlak mæXadoʃ forṣəthom homma Xado
[cvc] cvc.cv cv.cv.cvc cv.cv.cvc cvc.cvc.cvc cvc.cv cv.cv

Figure 4: present part of informant 1's speech sample transcription and segmentation process.

Note that:

- **P1** — = the first silent pause.
- **P2** — = the second silent pause.
- **P3** — = the third silent pause & So on.
- **Between brackets [ ]** = filled pauses, hesitations and corrections. Like **P5 & P7**.

[16]For all the ten informants' transcription and segmentation processes see appendices.

https://phonetics-acoustics.blogspot.com

b) Speech rate SR measurements

The measurements of speech rate SR is calculated by dividing the number of all the produced syllables by the whole speaking time (which is 30 seconds for each speaker). For measuring the articulation rate AR, the articulation time should be counted first (the main time of articulated syllables without any pauses or hesitations).The researcher excluded all the pauses durations (through cutting all the pauses and hesitations by Praat software) from the speech sample of each speaker. The whole duration that remained, after excluding all pauses and hesitation time, is the articulation time (the duration of the articulated syllables only).For example, see table 1.

c) Articulation rate AR measurements

The articulation rate AR is calculated by dividing the number of the articulated syllables (excluding any hesitations, corrections and repetitions) by the articulation time. For example see table 1.

Table 1: showing some measurements of informant's 1SR & AR.

Informant 1	
Gender	Male
Number of all pronounced syllables	181 (Syll.)
SR	6.033 (Syll. /Sec.)
AR	8.012 (Syll. /Sec.)
AR time	21.59 (Sec.)

Note that:

- **SR** = $\frac{\text{The total number of all pronounced syllables}}{\text{30 seconds}}$
- **AR** = $\frac{\text{The number of all pronounced syllables excluding all pauses and hesitations}}{\text{AR Time in seconds for each informant}}$

https://phonetics-acoustics.blogspot.com

d) Pauses calculations

All the pauses (silent and filled) that occurred in each informant's speech sample (which is 30 seconds) had been marked and numerated, each one of them, according to their occurrences to specify the number of pauses and their types for each informant. Then, the duration of each one of those pauses had been measured and each one's type had been identified. For example see table 2 which exposes a list of pauses of informant 1 with their number, types and durations[17].

Table 2: showing list of the distribution of all pauses with their types and durations in informant 1

Pause Number	Pause Type	Pause Duration (msec.)	Pause Number	Pause Type	Pause Duration (msec.)
P1	Silent	410	**P14**	Silent	390
P2	Silent	190	**P15**	Silent	170
P3	Silent	210	**P16**	Silent	330
P4	Silent	770	**P17**	Silent	360
P5 [Filled]	Prolongation	510	**P18**	Silent	440
P6	Silent	590	**P19**	Silent	360
P7 [Filled]	Repetition	200	**P20 [Filled]**	False start	485
P8 [Filled]	Retracting	207	**P21**	Silent	360
P9	Silent	560	**P22 [Filled]**	Repetition	208
P10	Silent	240	**P23 [Filled]**	Repetition	278
P11 [Filled]	False start	198	**P24**	Silent	230
P12	Silent	230	**P25**	Silent	140
P13	Silent	180			

[17] For each informant list of pauses with their numbers, types and durations see appendices.

https://phonetics-acoustics.blogspot.com

Also the percentage of all pauses and hesitations to the overall speech had been calculated through dividing the sum of all pauses durations (that occurred in the informant's speech sample) by the whole speech sample's duration (30 seconds for each informant). In addition; the researcher had calculated the degree of hesitancy which equals the proportion of filled pauses occurrences to all pauses that occurred in the whole speech sample (30 seconds) for each informant. See table 3

Table3: showing pauses' measurements of informant1 speech sample.

Informant 1	
The number of pauses (frequencies of occurrences)	25
Total pauses time	8.41 sec.
Percentage of all pause and hesitations to overall speech time	28.03 %
The degree of hesitancy to over all speech	28 %

Note that:

- **Total pauses duration** = Sum of each pause duration = p1 + p2 + p3 + p4 + p5 + ⋯
- **Percentage of all pauses and hesitations to overall speech =**

  $$\frac{\text{Total pauses durations in seconds}}{\text{30 seconds}} X\ 100$$

- **The degree of hesitancy**[18] **=** The proportion of filled pauses to all pauses

  $$= \frac{\text{Number of filled pauses}}{\text{Number of all pauses}} X\ 100$$

[18]According to Künzel (1997) the resultant proportion of this formula is considered an important forensic parameter because it tended to show large differences between speakers and small *within*-speaker differences.

https://phonetics-acoustics.blogspot.com

e) F_0 measurements

The following figure showing the fundamental frequency F_0measurements (mean pitch) for each speaker through Praat Voice Report (see Figure 5 for example).

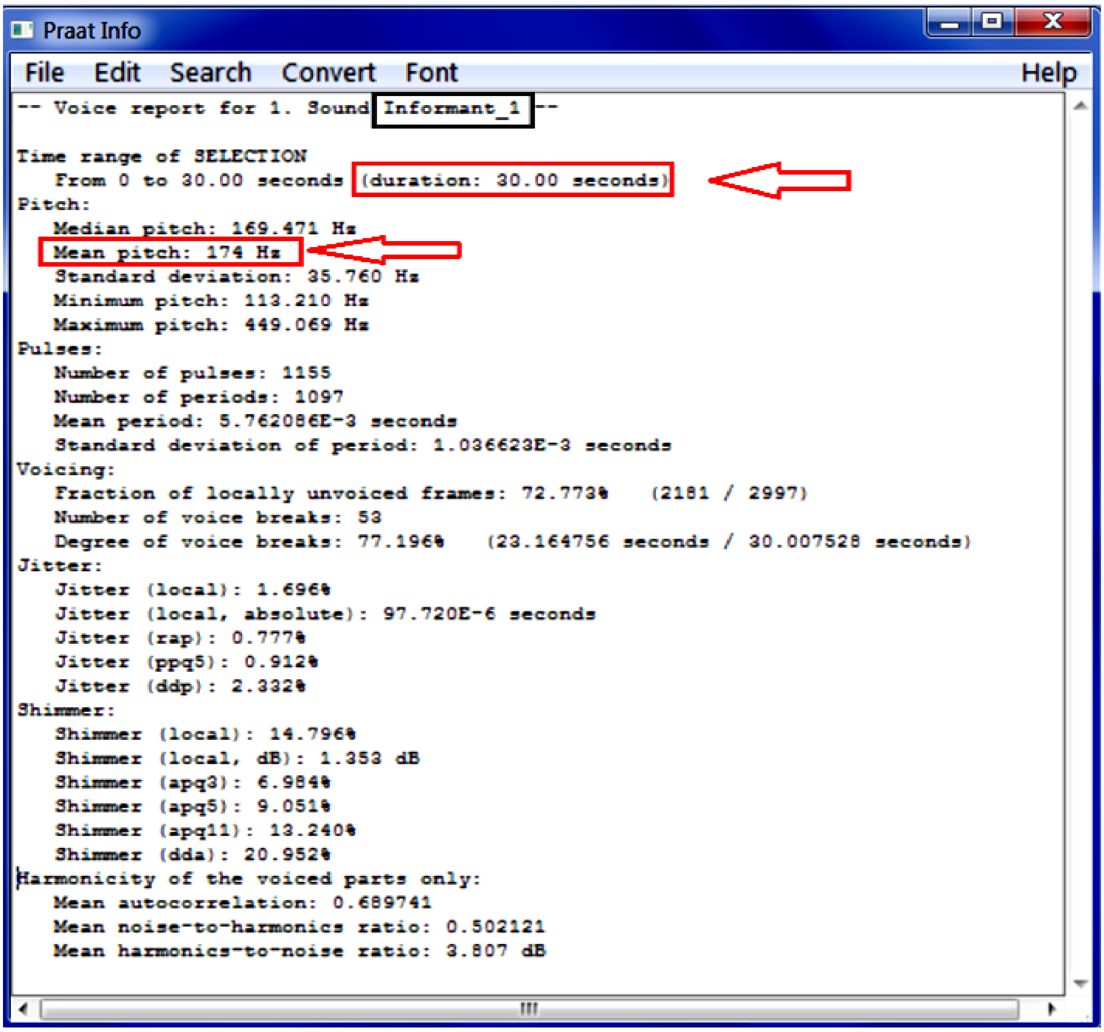


Figure 5: Voice Report of informant 1 from Praat Software.

f) Intensity measurements

The mean intensity for each speech sample is measured by Praat software through getting the mean intensity for each speaker's speech sample (see table 4).

https://phonetics-acoustics.blogspot.com

Table 4: showing a comprehensive exhibition of all the measurements for all the ten informants for 30 seconds for each informant (Speech Rate Time). The ten informants (5 males and 5 females) arranged one by one.

Informants / *Measurements*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Gender	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
F_0*(Hz)*	174	279	187	212	114	245	120	245	174	247
Intensity *(dB)*	60	82	75	77	69	71	73	71	77	79
Number of all pronounced syllables *(syll.)*	181	207	216	192	167	173	201	194	155	166
Speech Rate SR *(Syll. /Sec.)*	6.0	6.9	7.2	6.4	5.6	5.8	6.7	6.5	5.2	5.5
Articulation Rate AR *(Syll. /Sec.)*	8.012	8.459	8.520	7.397	6.731	6.921	7.723	6.95	6.015	5.889
Articulation Time *(Sec.)*	21.59	21.87	23.24	23.93	22.88	23.55	24.99	26.90	24.77	26.49
The number of pauses *(frequencies of occurrences)*	25	22	17	22	18	15	11	10	16	14
Total pauses time *(Sec.)*	8.41	8.13	6.46	6.07	7.12	6.45	5.01	3.1	5.23	3.51
Percentage of all pauses and hesitations to the overall speech %	28.03	27.1	21.5	20.23	23.73	21.5	16.7	10.33	17.43	11.7
The degree of hesitancy %	28	54.5	23.5	50	66.7	26.7	27.3	50	31.3	35.7

https://phonetics-acoustics.blogspot.com

4.2. Statistical representation of all the results

a) Speech rate SR and articulation rate AR values for all speakers.

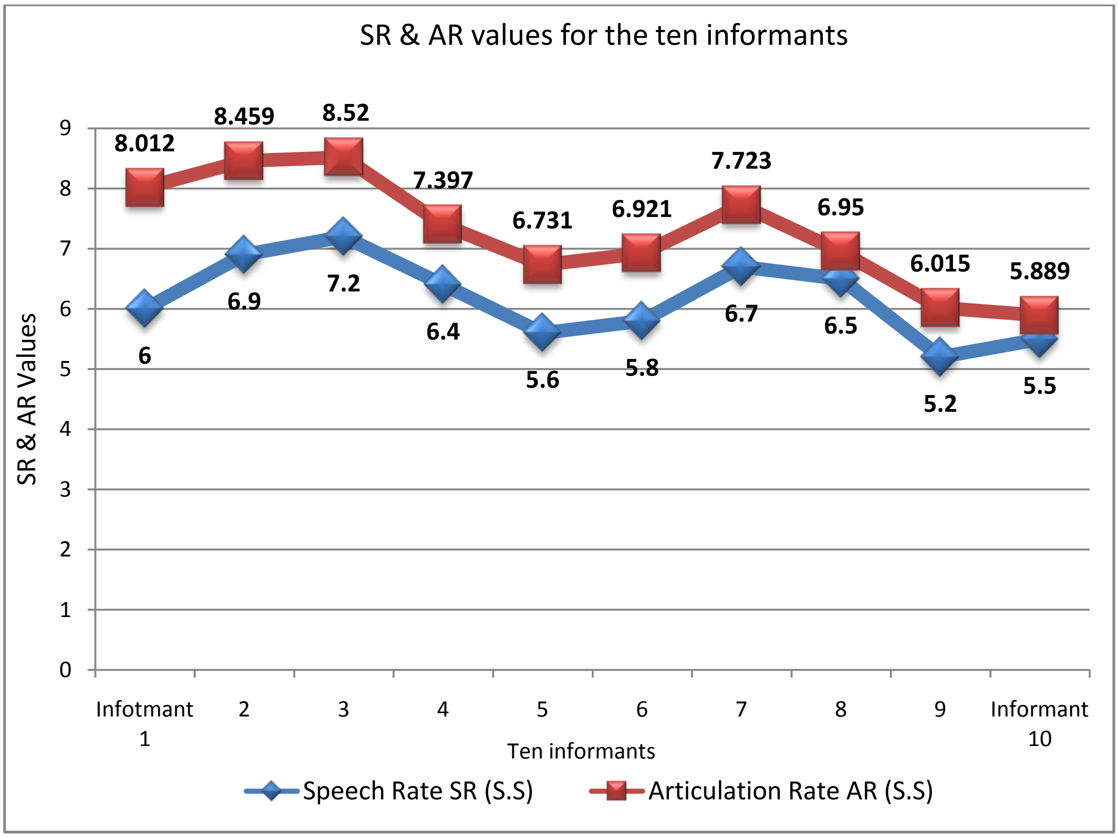


Figure 6: line- chart showing the values of the speech rate SR and the articulation rate AR for all the ten unknown informants.

Figure 6 represented speech rate SR and articulation rate AR values for all the ten informants. According to the preceding figure, we can deduce the following:

With respect to ***the fastest speaker***; informant 3 (male) is the fastest speaker with the highest SR= 7.2 S.S. and AR= 8.520 S.S. He pronounced the largest number of syllables in 30 seconds (216 syllables); he also has the highest F_0 between male speakers (187 Hz).

The second fastest speaker is informant 2 (female) who has SR = 6.9 S.S. and AR= 8.459 S.S. She has the second place in rate according to the number of the pronounced syllables (207 syllables) in 30 seconds. She has the highest F_0 between females (279 Hz); and she also has the highest intensity (82 dB) across all speakers.

The third fastest is Informant 7 (male) who has the third place according to the number of the pronounced syllables (201 syllables) and consequently in speech rate SR = 6.7 S.S.; but not in articulation rate AR = 7.723 S.S.; and he also has a relatively low F_0 (120 Hz) across male speakers but not the least one (See figure 6 and table 4).

https://phonetics-acoustics.blogspot.com

Regarding ***the slowest speaker speech rate SR***; informant 9 (male) is the slowest speaker with SR= 5.2 S.S. and he produced the minimum number of syllables in 30 seconds (155 syllables).

Informant 10 (female) takes the second place at the slow speech rate SR= 5.5 S.S. and the number of pronounced syllables = 166 syllables. She also has a relatively high F_0 (247 Hz) between female speakers; and also, she has a relatively high intensity (79 dB) across all speakers. Informant 10 also has a few numbers of pauses; only 14 pauses overall her total speech sample. Moreover, their total duration is very short (3.51 Sec. of total 30 seconds)

According to ***the articulation rate AR of the slowest speaker***; it's noticed that, informant 10 (female) has the slowest articulation rate AR = 5.889 S.S. followed by informant 9 (male) AR = 6.015 S.S. which indicates that the arrangement of informant 9 and informant 10 is reversed only in their articulation rates AR results.

Informant 5 (male) has the third place at the slow speech rate and articulation rate too; SR = 5.6 S.S., AR = 6.731 S.S., and he pronounced 167 syllables in 30 seconds. He also has the lowest F_0 (114 Hz) between male speakers and he has the second small intensity degree (69 dB) but not the least. Moreover, he has the highest degree of hesitancy (66.7 %). See figure 6 and table 4.

- **Correlation between speech rate SR and articulation rate AR**

Acoustically, there are many more acoustic cues which modify and determine the perceived speaking tempo. According to the results of the present experiment, the fastest speaker in speech rate SR has the fastest articulation rate AR as well (see Figure 7 and 8). But the slowest speaker in speech rate SR is not having the slowest articulation rate AR (see Figure 7 and 8) which indicates that the articulation rate AR has a relatively significant effect in identifying the fastest or the slowest speaker. Those results imply that speech rate SR is more prominent in identifying speakers with the fastest speaking tempo as well as slowest speaking tempo. Moreover, the number of the pronounced syllables which depend on the velocity of the speech organs is the main factor that directly influence the speech rate SR specifically.

https://phonetics-acoustics.blogspot.com

https://phonetics-acoustics.blogspot.com

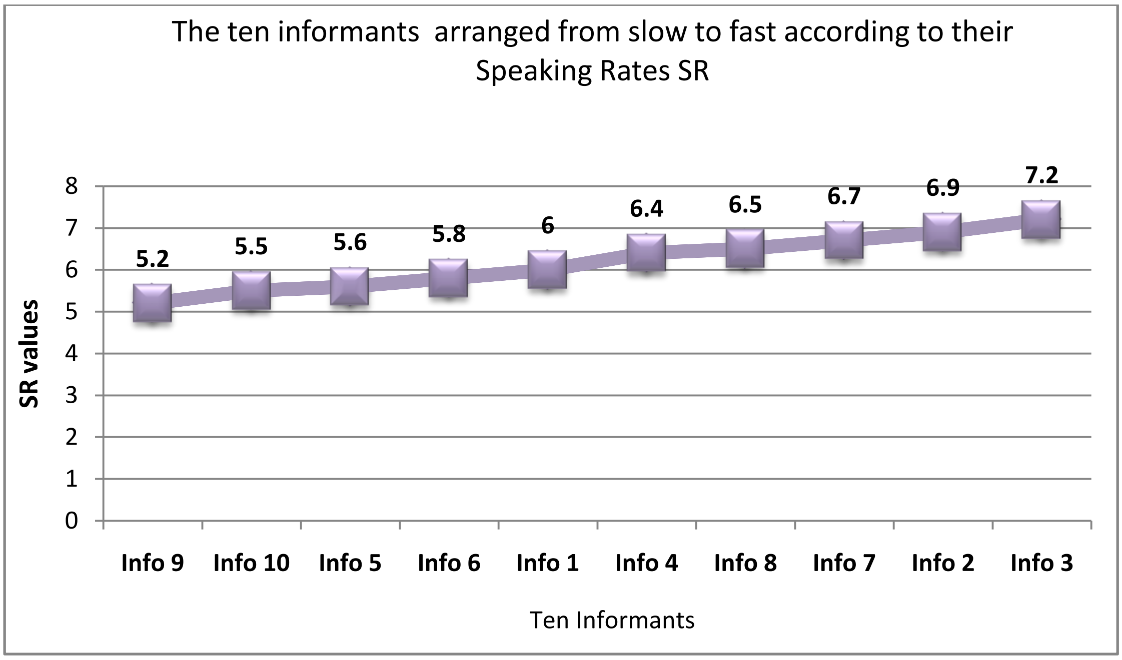


Figure 7 : Line chart used to visualize the speech rate SR differences across the ten informants who organized according to their speaking rates SR from the slowest to the fastest (from left to right).

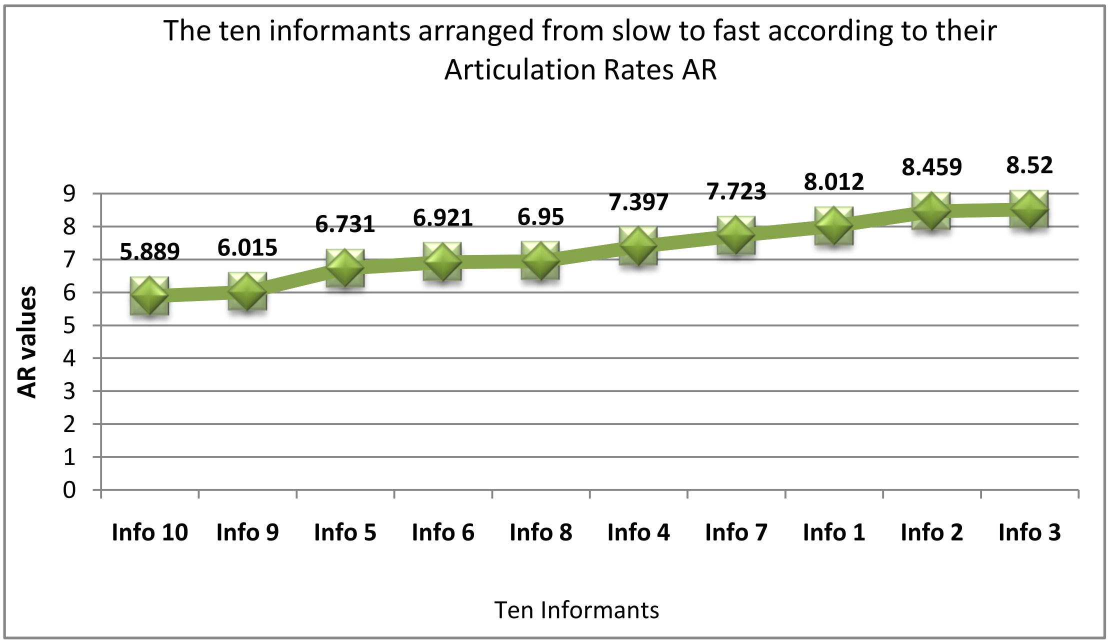


Figure 8 : Line chart used to visualize the articulation rate AR differences across the ten informants who organized according to their articulating rate AR from the slowest to the fastest (from left to right).

Romito, Lio & Galata (2005) states that the articulation rate AR depends on the intrinsic duration of the various phones, the rate of the articulation movements and on the rules of co articulation, but speech rate SR depends on speaker-specific features and on the communicative situation.

Koreman (2006) investigates the role of the articulation rate AR in distinguishing fast and slow rates. Koreman finds out that the intended and realized phone rates alone (Articulation Rate AR) can't explain the perceived rates of the speakers, which indicates that there are other factors than the articulation rate AR, that determine the perceived speaking tempo such as pauses and hesitations (which are excluded from his data). According to his results, we can't assume a direct relationship between articulation rate AR and the perceived rate.

Moreover, Laver (1994), reports that Goldman - Eisler (1968) shows, experimentally, that the speech rate SR and the articulation rate AR have no significant correlation. But, this doesn't exclude the possibility that articulation rate AR is a concomitant of speech rate SR.

On the contrary, Gold (2012), investigates the role of articulation rate AR as a discriminant in forensic speaker comparisons. Gold shows experimentally that the articulation rate AR performs much better as a discriminated parameter *within* the same speaker comparisons (indicating different styles from the same speaker) than the different speaker comparisons. He manifested that articulation rate AR is still considered in forensic speaker comparisons in conjunction with other speech parameters. This is based on Rose (2006) who points out that "not all speakers differ from each other in the same." Therefore, there will be individuals where articulation rate AR is a highly discriminated parameter for them. Also Jessen (2008) assures that when articulation rate is taken into account, considering only the fluent speech and all pauses ignored, the *between* speaker variation is relatively greater, and the *within* speaker variation seems to be smaller in articulation rate than syllable rate (Goldman Eisler, 1968; and Künzel, 1997), which makes articulation rate AR the more promising measure for forensic purposes.

The speech rate SR (slow or fast) seems to depend on the velocity and the flexibility of the articulators involved in the speech production. As a matter of fact; Dellwo, Kolly & Leemann (2012) claims that the movements of the speech organs (jaw, lip and tongue) are considered to be idiosyncratic, as well as the ways to control them. Also, Tillmann & Pfitzinger (2003), in research about the relation between kinematics and the speech rate, proved that the speech rate SR is mainly related to the velocity and the acceleration of the articulators which is primarily responsible for the co articulation processes. The research also assures, through using visual evidences, that when the whole range of movements remain almost constant, corresponds to large amount of speech rate variation presented in the utterance.

https://phonetics-acoustics.blogspot.com

b) Mean Intensity.

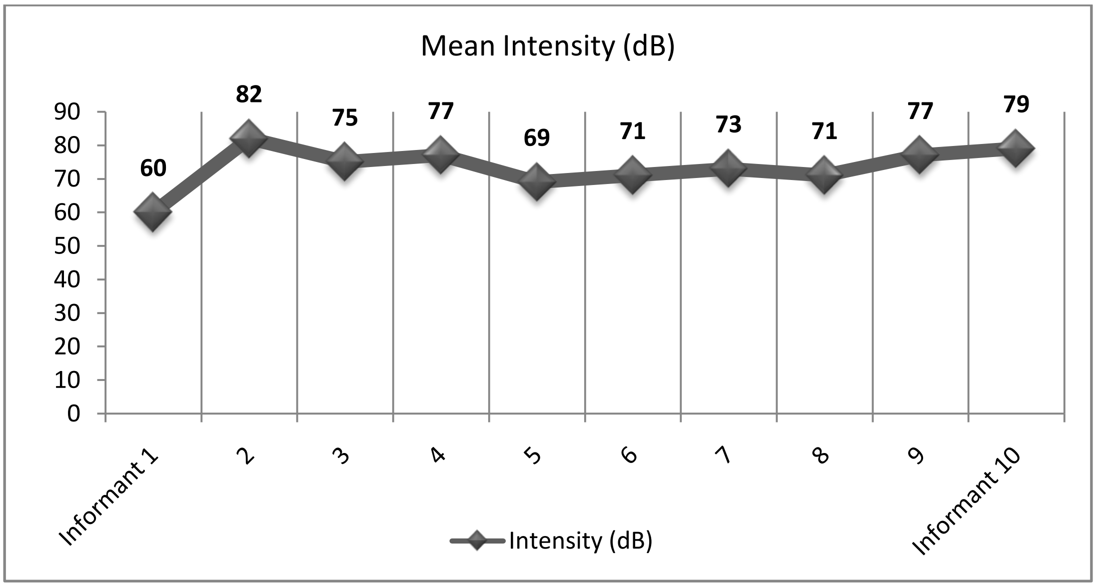


Figure 9 : Line chart used to visualize the mean intensity of the total speech duration for the ten informants.

According to the preceding figure (figure 9), regarding high intensity degrees, informant 2 (female) recorded the highest degree of intensity = 82 dB; followed by informant 10 (female) who recorded 79 dB.

Observe that, informant 2 who has the highest intensity degree, is the second fast speaker. And informant 10 who has the second high degree of intensity is the second slow speaker.

Regarding low intensity degrees, informant 1 (male) recorded the lowest intensity degree = 60 dB; followed by informant 5 (male) who recorded 69 dB. Perceptually, 41% of the listeners expected informant 1 to be the slowest speaker (see the perceptual test results; section 4.3).

https://phonetics-acoustics.blogspot.com

c) Percentage of pauses to overall speech sample for each speaker.

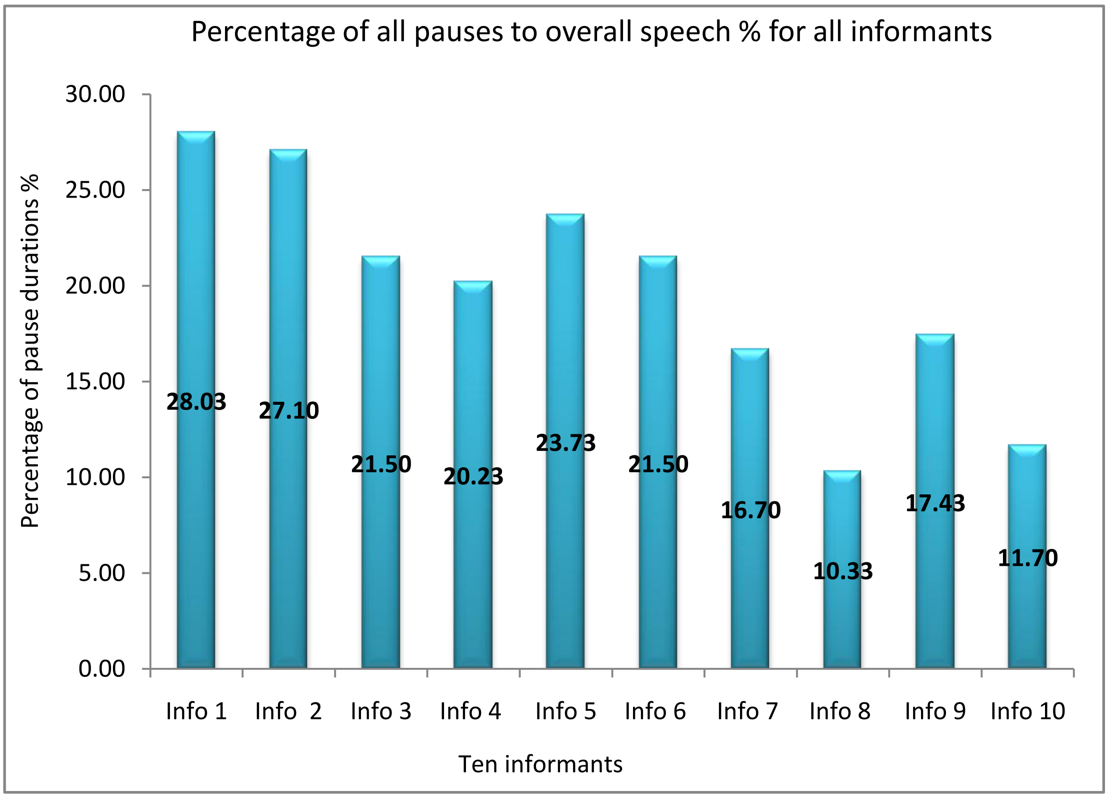


Figure 10: column - chart showing the percentage of all pauses and hesitations to the overall speech sample for each speaker.

Primarily the percentage of all pauses to the overall speech sample depends on both the number of pauses (pauses' frequencies of occurrences) as well as their durations. For more information about the number of pauses and their durations see appendices.

According to the preceding figure (see Figure 10), respecting ***the highest percentage of all pauses to overall speech sample***, informant 1 has the highest percentage of pauses (28 % of his speech sample consists of pauses); followed by informant 2 (27.10 % of her speech sample consists of pauses); followed by informant 5 (23.73 % of his speech sample consists of pauses). See appendices for more information about pauses.

Regarding ***the lowest percentage of all pauses to the overall speech sample***, the preceding column-chart (see Figure 10) indicates that informant 8 has the lowest percentage of pauses (only 10.33 % of her speech sample consists of pauses); followed by informant 10 (11.7 % of her speech sample consists of pauses); followed by informant 7 (16.7 % of his speech sample consists of pauses). See appendices for more information about pauses.

https://phonetics-acoustics.blogspot.com

Hence, informant 8 and informant 10 should be expected to be the fastest speakers according to their percentage of all pauses to the overall speech sample because they have the lowest percentages; and reversely, informant 1 and informant 2 should be expected to be the slowest speakers because they have the highest percentages of pauses to overall speech. But this is not true acoustically, which implies that the percentage of all pauses to the overall speech sample has no direct influence on the speech rate and they cannot determine or modify the speech rate. The results of the present experiment indicate that the percentage of all pauses to overall speech sample plays a double-edged role:

The first role of the percentage of all pauses is on the acoustical level, they don't have any obvious effectiveness on the speech rate SR; i.e., there is no remarkable correlation between the speech rate SR and pause (the fastest speaking rate SR must not have the minimum percentage of pauses and vice versa). However, Laver (1994), reports that Goldman - Eisler (1968) shows, experimentally, that speech rate SR may be positively correlated with the percentage of pauses and durations in the speech sample. Moreover, Vaane (1982) states that the most languages' variations in speech rate SR are mainly due to variations in the durations of pauses. However, in light of the results of the present experiment, it does not appear to be a verity. The results of the present experiment also contradict with Fashal (1991) which, experimentally, studies the tempo of reading texts of news-bulletins in Egyptian broad casting. The results conclude that the number and the duration of pauses is an effective factor in changing the rate of speech i.e., decreasing the number and the duration of the pauses increases the speech tempo.

The second role of the percentage of all pauses is on the perceptual level, large percentage of pauses durations is considered one of the most important factors that influence the listeners' perceptions; in particular, the perception of slow speaking tempo. Conversely, small percentage of pauses durations does not mean fast speaking tempo for listeners' perceptions. That's confirmed by Koreman (2006) who states that; it is possible that; pausing or dis-fluencies determine the perceived speaking tempo.

https://phonetics-acoustics.blogspot.com

d) The degree of hesitancy for all speakers.

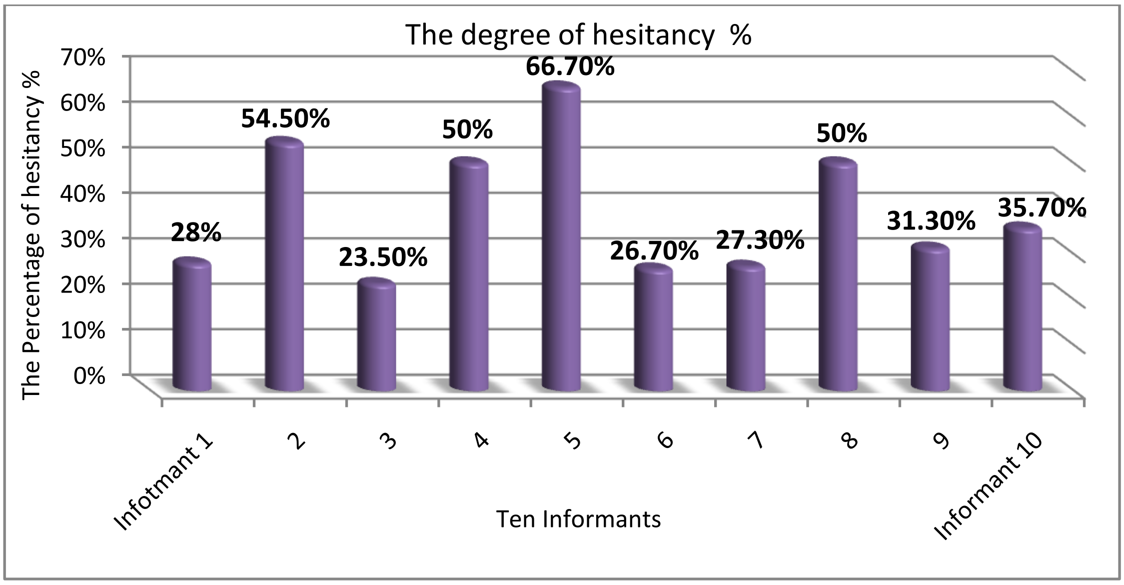


Figure 11: cone - chart showing the degree of hesitancy % for each informant.

Rose (2002) reports that according to Künzel (1997), the degree of hesitancy for each informant shows the proportion of filled pauses to all pauses for the overall speech sample to indicate large differences *between* speakers (intra speaker variation) and relatively small differences *within* speaker (inter speaker variation).

The preceding cone-chart (see Figure 11) indicates that informant 5 (who is arranged as the third slow speaker according to his speaking rate and he has the lowest F_0 between male speakers) has the highest degree of hesitancy (66.7 %), which may indicate that the high degree of hesitancy may negatively affect the perceived speaking rate. In other words; high degree of hesitancy may be considered as a sign of slow speaking rate. To confirm this, we need more experimental research.

Informant 2 has the second high degree of hesitancy = 54.5 %. (Informant 2 is arranged as the second fast speaker according to her speaking rate).

Figure 11 also indicates that informant 3 (who is the fastest speaker according to his speaking rate and who has the highest F_0 between male speakers) has the lowest degree of hesitancy (23.5 %).Followed by informant 6 (26.7 %); followed by informant 7(27.3 %) who is arranged as the third fast speaker according to his speech rate.

Regarding the results of the present experiment; the degree of hesitancy seems to have an inverse relation with speaking tempo particularly at fast speaking tempo. In other words; the fastest speaker (according to speech rate SR and articulation rate AR) has the least degree of hesitancy. And this relation is compatible with only the fast speaking rate.

https://phonetics-acoustics.blogspot.com

4.3. Perceptual test results and their statistical representations.

Sixty naïve listeners are involved in the present experiment for doing the perceptual test and selecting the fastest speaker and the slowest speaker too. The following figure (Figure 12) showing the distribution of all the listeners' selections percentages for both the fastest speaker and the slowest speaker as well. Through glancing over Figure 12, it's noticed that, the percentages of listeners' selections are highly distributed across all the ten informants with varied degrees which reveal that there is no absolute agreement about a particular speaker whether the slowest or the fastest.

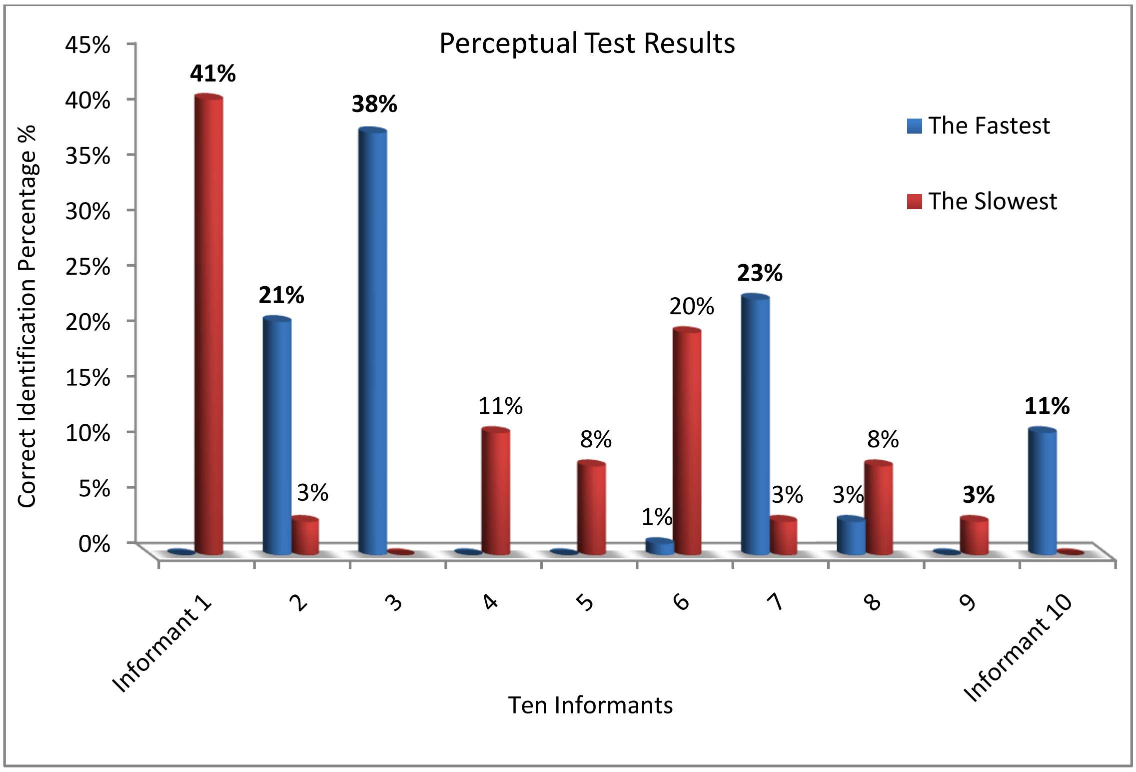


Figure 12: line- chart showing the distribution of all the listeners' selections (correct and false identifications) of both the fastest informant and the slowest informant depending only on their ears.

With respect to ***the fastest speaker identification***; 38 % of the listeners select informant 3 to be the fastest speaker (with the fastest speaking rate). And 23 % of the listeners select informant 7 to be the fastest speaker, followed by 21% of the listeners who select informant 2 to be the fastest speaker (see Figure 12).

On the other hand, the results of ***the slowest speaker identification***; showed that41% of the listeners expect informant 1 to be the slowest speaker. And 20 % of the listeners select informant 6 to be the slowest speaker. These results indicate that identifying the slowest speaker seems to be more problematic for the listeners (see Figure 12).

https://phonetics-acoustics.blogspot.com

- **Correlation between the acoustical results and the perceptual ones**

Hayward (2000) said that, "experiments in speech perception have commonly focused on a single acoustic cue and a single impressionistic-phonetic dimension. Nevertheless, it would be highly exceptional for a single cue to be solely responsible for signaling a simple phonetic contrast". For more illustration; if listeners are asked to classify stimuli which vary in two or more cues, there is typically a *trading relation* between the cues. Hayward (P. 118) adopted this trading relation hypothesis; explaining that this trading relation is used in the sense of *trade-off* which is used in everyday situations. In other words, for the listeners' perception while listening to a speech sample, some acoustic cues are more dominant or more prominent above some other cues.

Regarding ***the fastest speaker identification;*** the acoustical results are corresponded with the perceptual results indicating that informant 3 (male) is the fastest speaker. According to the perception of the listeners, 38 % of the listeners select informant 3 (male) to be the fastest speaker (see figure 12). According to the acoustical measurements, as well, informant 3 (male) pronounced the largest number of syllables in 30 seconds (216syllables); so, he has the highest SR = 7.2 S.S. and AR = 8.520 S.S.; moreover, he has the minimum degree of hesitancy 23 % and the highest F_0 between male speakers 187 Hz (see table 4).

Perceptually, 23 % of the listeners expected that informant 7 (male) is the fastest speaker. This proportion can't be easily ignored; informant 7 is arranged acoustically as the third fastest speaker according to the number of the pronounced syllables (201 syllables) and consequently in speech rate (SR = 6.7 S.S.). But informant 7 has a relatively small articulation rate AR = 7.723 S.S.; and a low F_0 (120 Hz) across male speakers. He also has a small number of pauses; only 11 pauses in his total speech with 27.3 % degree of hesitancy (see table 4 and appendix 7). By gathering all the preceding acoustical cues together; perceptually, informant7 is considered to be a fast speaker but not the fastest.

21 % of the listeners expected that informant 2 (female) is the fastest speaker. Informant 2 acoustically is organized as the second fastest speaker according to the number of the pronounced syllables. She had produced207 syllables in 30 seconds; and consequently, her speaking rate results (SR= 6.9 S.S. and AR= 8.459 S.S.). Moreover, she has 22 pauses in her total speech sample with relatively high degree of hesitancy 54.5 % (see table 4 and appendix 2).Informant 2 also has the highest F_0 between females (279 Hz); and she also has the highest mean intensity (82 dB) across all speakers. All the preceding cues together gave an impression of fast speaking tempo for the listeners. Perceptually, listeners' perceptions seem to be highly affected by her remarkable highest degrees in F_0 and

https://phonetics-acoustics.blogspot.com

intensity across all speakers more than the number of pauses and the degree of hesitancy.

11 % of the listeners also expected that informant 10 (female) is the fastest speaker. Whereas, she is arranged acoustically as the second slowest speaker according to her speech rate SR= 5.5 S.S. and she has the slowest articulation rate AR across all the involved speakers (AR = 5.889 S.S.). However, no one of the listeners selects her or expects her to be the slowest speaker (see figure 12).On the other hand, Informant 10 has a relatively high F_0 (247 Hz) between female speakers; and also, she has the second high intensity degree (79 dB) across all speakers. Informant 10 also has a few number of pauses; only 14 pauses overall her total speech sample; moreover, their total duration is very short (3.51 Sec. of total 30 seconds), see table 4. These preceding results indicated that, high $F_0$and high intensity; in addition to her few number of pauses and their short duration; all of these cues together may be responsible for perceiving fast speaking tempo more than the other acoustic cues.

With respect to ***the slowest speaker identification***, the perceptual results did not match with the acoustical results.

Perceptually, 41 % of the listeners expected that Informant 1 (male) is the slowest speaker. Acoustically, informant 1 has recorded the largest percentage of all pauses to the overall speech sample = 28.03 % (see figure 10). Informant 1 also has the largest number of pauses' frequency and duration; 25 pauses in his total speaking time; in 8.41 seconds from 30 seconds (see table 4 and appendix 1). Moreover; he has the minimum intensity degree (60 dB) across all speakers. The preceding acoustical cues imply that there is a correlation between the number and the duration of pauses with the perception of the listeners' particularly in determining the slow speaking rates. In other words; increasing the number and the duration of the pauses gives a perceptual impression of slow speaking rate. In addition to, the mean intensity seems to have a remarkable affect on the listeners' perception i.e., the lower the loudness, the slower the perceived speaking rate and vice versa (see chapter 2 section 2.3.3).

On the other hand; perceptually, only 3 % of the listeners select informant 9 to be the slowest speaker. According to the acoustical results informant 9 is the slowest speaker, because he pronounced the minimum number of syllables (155 syllables) in 30 seconds; and consequently; his speaking rate was the smallest (SR= 5.2 S.S.) and his articulation rate was relatively small (AR= 6.015 S.S.). Informant 9 has a relatively high F_0 between male speakers 174 Hz, in addition to a relatively high mean intensity 77 dB (see table 4). These preceding results indicated that; because of his relatively high F_0 and relatively high mean intensity only 3 % of the listeners select him to be the slowest speaker; and no one select him to be the fastest speaker as well.

https://phonetics-acoustics.blogspot.com

To summarize, from the preceding exhibition of the correlation between the acoustical results and the perceptual results, we can deduce the following: F0, intensity and pauses' numbers and durations have significance effect on the perceived speaking tempo. And further, those acoustical cues are used as perceptual parameters that almost all of the listeners relay on, to identify the slowest and the fastest speaking rates.

There are so many parameters that influence the speaking tempo and its' perception. Naïve listeners may focus on some of these cues more than others; which give the possibility of false identification to occur. Particularly, in identifying the slowest speaking tempo; most of naïve listeners clearly relate the slow speaking tempo directly with the pauses frequencies and durations which confuse most of them in identifying the slowest speaking tempo. Whereas the phonetician expert must find all the sources and types of these cues and combine them together for identifying the identity of the unknown speaker in the forensic case work.

https://phonetics-acoustics.blogspot.com

CHAPTER 5

5.1. Conclusion

It has become clear over the years that speaker identification should not be confined to a single method, but a variety of different methods should be used. To find out the most effective acoustical cues which affect the rate of speech and its perception, an acoustic – perceptual line must be drawn. Clearly, the acoustical results should explain and clarify the perceptual ones. According to the acoustical results and the perceptual results of the present study, we deduced the following conclusions:

- With respect to **speech rate SR** and **Articulation rate AR:** Speech rate SR is more prominent in identifying the unknown speaker. However, this does not mean to exclude the articulation rate AR.

  On the perceptual level; in identifying the rate of speech, listeners are listening to the speech rate SR (with all pauses i.e. filled, silent and hesitations). They are not listening to the articulation rate AR only; therefore, articulation rate AR should be combined with other acoustic cues in order to reflect perceptually a fast or a slow speaking tempo.

  On the acoustical level; articulation rate AR is distinctive for some speakers because it referred directly to the pure pronounced syllables which related primarily to the velocity of the speech organs of the speaker (whoever male or female). Substantially, velocity of the speech organs is an effective factor in modifying the tempo of speaking. As a result of that; speech rate SR should be combined with the articulation rate AR in order to help in identifying the identity of the unknown speaker under the forensic circumstances.

- **The degree of hesitancy:** acoustically; it is considered as a remarkable factor for *the fastest speaking tempo*. In other words, according to the results of the present experiment; the fastest speaker in speaking tempo has the least degree of hesitancy. But the highest degree of hesitancy does not take place at the slowest speaker's speaking tempo.

- **F_0:** it is an important acoustic cue in identifying the speaker's speaking rate acoustically and perceptually as well. The importance of F_0, according to our results, appeared in identifying *the fastest speaking tempo*; but not in identifying the slowest speaking tempo. High F0 (for male or female speaker) indicated fast speaking tempo perceptually as well as acoustically.

https://phonetics-acoustics.blogspot.com

- **Mean intensity:** *on the perceptual level*, mean intensity is a remarkable cue for listeners' perception in identifying the rate of speaking of the speaker (whether the slowest or the fastest). In other words, high intensity indicated fast speaking tempo; and low intensity indicated slow speaking tempo. Whereas, *on the acoustical level*, and according to our results, mean intensity results are inconsistent with speaking tempo results.

- **The percentage of all pauses**, **pauses' numbers** and **pauses' durations** have a double aged role.
  On the perceptual level; naïve listeners have a great ability in detecting different speaking tempi. However, particularly in identifying the slowest speaking tempo, there is some inconsistency between the acoustical results and the listeners' selections. Seemingly, listeners received the high percentage of all pauses as a sign of low speaking tempo that may confuse them in identifying the slowest speaking tempo, unlike identifying the fastest speaking tempo.
  More specifically, informant 1 has the lowest mean intensity and the highest percentage of all pauses. Therefore; 41 % of the naïve listeners expected him to be the speaker with the slowest speaking tempo. Sequentially as a result, on the perceptual level and according to the results of the present experiment, low loudness and high percentage of all pauses, when occurred together in that way, seem to have a great influence on the listeners' perception in identifying the slowest speaking tempo particularly.

  On the acoustical level; according to the results of this study, the percentage of all pauses, pauses' numbers and pauses' durations did not show significance role in modifying the rate of speech. For more illustration, the fastest speaker with the fastest speaking rate does not have the least degrees of anyone of these cues.

https://phonetics-acoustics.blogspot.com

Summary

Introduction

Speaker Identification is the task of deciding and determining a given sample of speech (uttered by unknown speaker), who among many candidate speakers said it. The unknown speaker is defined as the speaker hose model best matches the given utterance (Furui 2008).

Forensic Speaker Identification FSI is considered as one of the most significant practical applications of speaker identification. FSI is defined as the most central aspect of forensic phonetics and acoustics which mainly concerned with solving problems related to identification of the unknown speaker in criminal investigation to identify suspects who were heard but not seen committing a crime including; murder, blackmail threats, ransom calls, kidnapping, political corruption, bomb threats, terrorist activities, etc. (Singh, Khan& Shree, 2012; Jessen, 2008; Nolan, McDougall, DeJong & Hudson, 2006; Lindh, 2004; Eriksson, 2005; Rose, 2002; and Nolan, 2005).

There will be always differences (which are always audible, measurable and quantifiable) between speech samples, even if they come from the same speaker. This is due to two kinds of variability: ***1) organic vs. phonetic variability***, and ***2) between speaker vs. within speaker variability***. Consequently, the main task of Forensic Speaker Identification FSI is to find all the sources of variability in order to make a clear distinction for the correct evaluation.

For speaker identification in forensic situation as evidence in the court, there are four main phonetic/acoustic parameters depending on the speaker through them he / she can be discriminated and identified:

1. The Fundamental Frequency F_0.
2. The formants frequencies of the vowels.
3. The resonance of the nasal consonants.
4. Tempo of speaking.

The fourth parameter is our concern here, which is a multidimensional phenomenon and revealing the temporal aspects of the speech.

Tempo of speaking is one of the prosodic cues which considered as non-linguistic factor that signaling ***paralinguistic*** information (about the situation and the inner state of the speaker's attitudinal or emotional state) and also ***extralinguistic*** information (about the speaker's identity, personality and individuality) (Trouvain, 2003; and Rose, 2002). Tempo of speaking has significant importance in Forensic Speaker Identification FSI Demenko (2000) because it is:

1. Carrying the individual-identifying information about the speaker.
2. Affected by the individuals variations in speaking.
3. Not affected by the frequency characteristics of the transmission systems and at the level at which the speaker talks.
4. Not easy to imitate or disguise.
5. Not controlled by the speaker.

Tempo of speaking can be exhibited by two methods, one is ***Speech Rate (SR)***, and another is ***Articulation Rate (AR)***. Both of SR and AR can be defined as "the number of syllables per second". The biggest difference between SR and AR is that

https://phonetics-acoustics.blogspot.com

the SR includes pause intervals but the AR does not (Gold, 2012; Koreman, 2006; and Fashal, 1991).

The fundamental goal of this study, on the acoustical level, is to prove acoustically that every speaker has a significant speech rate SR and articulation rate AR through which the unknown speaker can be discriminated and to investigate which of them (SR or AR) could be more benefit for identifying unknown speakers and to what extent. Also, the present study is essentially concerned, on the perceptual level, with listeners' perceptual abilities in perceiving and differentiating different speaking tempi for identifying unknown speakers in order to utilize this exceptional ability in forensic speaker identification FSI.

Methodology

The experiment includes 10 unknown speakers (5 females and 5 males) of colloquial Arabic language, aged between 19 and 40 years old. Natural spontaneous speaking style is elicited for 30 seconds for each speaker trying to avoid the effect of any stress or the domination of any specific emotion. All data are transcribed and segmented manually by the researcher with the aid of Praat software (www.praat.org) as a listening tool. Then the SR, AR, F_0, mean intensity, the degree of hesitancy, and number of pauses, pauses' durations and the percentage of all pauses to the overall speech time are calculated for each speaker.

Sixty listeners of university students, aged between 17 and 25 years old, are involved in the perceptual test. They are asked to listen carefully to the 10 unknown speakers three times at most, in order to select **the fastest** speaker and **the slowest** one too by marking (√) in front of his or her title in the received listening sheet.

Results

The results indicate that:

1. Acoustically and perceptually, SR is most powerful in identifying unknown speakers than AR. But this does not mean to exclude the articulation rate AR.
2. The percentage of all pauses plays a double-edged role. On the perceptual level, large percentage of pauses durations considered one of the most important factor that influencing the listeners' perception of the rate of speech. Whereas, on the acoustical level, they don't have any obvious effectiveness on modifying the rate of speaking.
3. The degree of hesitancy, acoustically, it is considered as a remarkable factor for the fastest speaking tempo. But not in identifying the slowest speaking tempo.
4. F0is an important acoustic cue in identifying the speaker's speaking rate acoustically and perceptually as well. High F0 (for male or female speaker) indicated fast speaking tempo.
5. Mean intensity, perceptually, is a remarkable cue for listeners' perception in identifying the rate of speaking of the speaker (whether the slowest or the fastest). High intensity indicated fast speaking tempo; and low intensity indicated slow speaking tempo.

https://phonetics-acoustics.blogspot.com

6. References

1- Adams, C., & Munro, R. (1978) In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. Phonetica, 35, 125-156.

2- Ainsworth, A.W. (1988) Some approaches to automatic speech recognition. *The handbook of phonetic sciences*, ed. by William J. Hardcastle and John Laver Ch. 24. Page 721 &722. Oxford, UK: Blackwell.

3- Alexander, A., Dessimoz, D., Botti, F. and Drygajlo, A. (2005) Aural and automatic forensic speaker recognition in mismatched conditions. *International Journal of Speech, Language and the Law* 12.214–34.

4- Allen, J. Sean, Joanne L. Miller, and Desteno, D. (2003) Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 113.544–52.

5- Amino, K., & Arai, T. (2009) Effects of linguistic contents on perceptual speaker identification: Comparison of familiar and unknown speaker identifications. *Acoustical Science and Technology*, *30*(2), 89-99.

6- Arvaniti, A. (1999) Effects of speaking rate on the timing of single and geminate sonorants. *Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 1: 599-602. San Francisco.

7- Battaner, E., Gil, J., Marrero, V., Llisterri, J., Carbó, C., Machuca, M. J., De La Mota, C., Ríos, A. (2003) VILE: Acoustic Study of Inter and Intra Speaker Variation in Spanish. Barcelona: SEAF, *Sociedad Española de Acústica Forense*. pp. 59-70. ISBN: 84-86108-32-2;978-84-86108-32-8. Website:http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_SEAF03_Engl.pdf.

8- Beck, J. M. (1997) Organic variation of the vocal apparatus. *The Handbook of Phonetic Sciences*, Oxford: Blackwell. Ed: (Hardcastle and Laver) .PP. 256 –297.

9- Beckman, M. E. (1989) Timing models for prosody and cross-word coarticulation in connected speech. Proceedings of the Second DARPA Speech Recognition Work-shop, 12-21. San Francisco: Morgan Kaufmann.

10- Berry, J. and Weismerb, G. (2013) Speaking rate effects on locus equation slope. Published in final edited form as: *Journal of Phonetics*. 2013 November; 41(6): 468–478. doi:10.1016/j.wocn.2013.09.002.

11- Bolt, R. H., Cooper, F. S., David, E. E., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1969) Speaker Identification By Speech Spectrograms. *Science*, pages 338-343.

12- Bolt, R. H., Cooper, F. S., David, E. E., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1970) Speaker Identification By Speech Spectrograms: A Scientist's View of Its Reliability for Legal Purposes. *J. Acoustic. Soc. Am.*, 47(2):597-603.

https://phonetics-acoustics.blogspot.com

13- Bolt, R. H., Cooper, F. S., David, E. E., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1973) Speaker Identification By Speech Spectrograms: Some Further Observations. *J. Acoustic. Soc. Am.*, 54(2): pages 531-534.

14- Bona, J., Graczi, T.E., Marko, A. (2008) Coarticulation rules and speaking style dependency. in: Sock Rudolph, Susanne Fuchs, Yves Laprie (ed.) International Seminar on Speech Production, Strasbourg: INRIA, str. 245-248

15- Bonastre, J. F., Bimbot, F., Boë, L. J., Campbell, J. P., Reynolds, D. A., and Magrin, C. I. (2003) Person Authentication by Voice: A Need for Caution. In Eurospeech 2003, Interspeech 2003. *Proceedings of the 8th European conference on speech communication and technology*. pp. 33-6. Geneva, Switzerland. Website:http://www.afcpparole.org/doc/AFCP_SpLC_HotTopicsEurospeech03_final.pdf

16- Broeders, A. P. A. (2001) Forensic Speech and Audio Analysis: Forensic linguistics 1998 to 2001. A review. *In 13th INTERPOL Forensic Sciences Symposium. Lyon, France*, October 16-19, 2001. Dept. of Handwriting, Speech and Document Examination. Netherlands Forensic Institute. Ministry of Justice, TheNetherlands.Website:https://www.interpol.int/Public/Forensic/IFSS/meeting13/Reviews/ForensicLinguistics.pdf

17- Busso, C., Lee, S., and Narayanan, S. (2009) Analysis of emotionally salient aspects of fundamental frequency for emotion detection. IEEE Transactions on Audio, *Speech, and Language Processing*, 17(4), 582-596.

18- Butcher, A. (1996) Getting the Voice Line-Up Right: Analysis of a Multiple Auditory Confrontation. In: eds: (P. McCormack & A. Russell). *Proceedings of the 6th Australian International Conference on Speech Science and Technology*. Canberra*: Australian Speech Science and Technology Association*, 97-102. butcher96.assta.pdf

19- Butcher, A.R. (2002) Forensic Phonetics: Issues in speaker identification evidence. Proceeding of the in augural *International Conference of the Institute of Forensic Studies*. Italy. pp. 3-5.

20- Cain, S., Smrkovski, L. and Wilson, M. (1990) Voiceprint Identification - Money Laundering and Narcotics Update, Department of Justice, 1988, and the Legal Investigator. NDAA Bulletin December 1993.

21- Cao, H. and Wang, Y. (2011) A Forensic Aspect of Articulation Rate Variation in Chinese. Hong Kong, 17-21 August 2011. Key Laboratory of Evidence Science (China University of Political Science and Law), Ministry of Education, China; Department of Chinese Language and Literature, Peking University, Beijing, China; Center of Criminal Technology, Public Security Bureau of Guangdong Province, China caohonglin@pku.edu.cn; wangyingli776@sina.com

22- Caspers, J. and Van Heuven, V.J. (1991) Phonetic and linguistic aspects of pitch movements in fast speech in Dutch. Proc. ICPhS Aix-en-Provence (5), pp. 174-177.

23- Cicres, J. (2011) Acoustic Discriminant Analysis of Full Pauses for Voice Comparison. Comunicación presentada en IAFPA 2011, Vienna.

https://phonetics-acoustics.blogspot.com

24- Clark J. & Foulkes P. (2007) Identification of Voices in Electronically Disguised Speech. *The International Journal of Speech, Language and the Law*, vol. 14 No 2 (2007) website: http://www.equinoxjournals.com/ojs/index.php/IJSLL

25- Clark, J. & Foulkes, P. (2006) Identification of Familiar Voices in Disguised Speech. Proceedings, IAFPA 2006.

26- Clopper, C. G. and Smiljanic, R. (2011) Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39. 237 – 245.

27- Dankovicová, J. (1999) Articulation rate variation within the intonation phrase in Czech and English. In *Proceedings of The Xivth International Congress Of Phonetic Sciences*, San Francisco. University College London, UK.

28- De Jong, G. (2007) Forensic Speaker Recognition. University of Cambridge.

29- Deese, J. (1984) Thought into Speech: The Psychology of Language. Prentice-Hall: Englewood Cliffs.

30- Dellow, V., Kolly M. J., and Leemann A. (2012) Speaker identification based on temporal information: a forensic phonetic study of speech rhythm and timing in the Zurich variety of Swiss German. Presented at IAFPA 2012, Santander/Spain, volker.dellwo@uzh.ch.

31- Dellwo, V. (2003) Forensic Phonetics. Phonetics and linguistics, *speech science*.

32- Dellwo, V. and Koreman, J. (2008) How speaker idiosyncratic is measurable speech rhythm? Abstract presented at the annual IAFPA meeting 2008, Lausanne/Switzerland. Attached to this application as IAFPA2008_DellwoKoreman.pdf.

33- Dellwo, V., Leemann, A., and Kolly, M. J. (2012) Speaker idiosyncratic rhythmic features in the speech signal. Proceedings of Interspeech 2012, 9.-13.9.2012, Portland (OR), USA.

34- Dellwo, V., Ramyead, S. y.,and Dankovicova, J. (2009) The influence of voice disguise on temporal characteristics of speech. Comunicación presentada en el Congreso de la IAFPA2009. Cambridge.

35- Demenko G. (2000) Analysis of supra-segmental features for speaker verification. Institute of linguistics, Adam Mickiewicz University, Poznan, Poland.

36- Doherty, E. H. (1975) *AN Evaluation of Selected Acoustic Parameters for Use in Speaker Identification*. PHD dissertation. University of Florida. Website: http://archive.org/stream/evaluationofsele00dohe/evaluationofsele00dohe_djvu.txt

37- El Nayal, H. (1991) *Determination of the basic phonetic parameters for speaker recognition*. M.A. thesis, Alexandria University, Faculty of Arts, Department of phonetics and linguistics.

38- Eriksson, A. (2005) Tutorial on forensic speech science. Part I: Forensic phonetics. In Inter-speech, Euro-speech 2005. Proceedings of the 9th *European conference on speech*

https://phonetics-acoustics.blogspot.com

communication and technology. Department of Linguistics, Gothenburg University, Gothenburg, Sweden 2005.
Website:http://www.york.ac.uk/media/languageandlinguistics/documents/currentstudents/Eriksson_tutorial_paper.pdf

39- Farnetani, E. (1990) Coarticulation and connected speech processes. *The Handbook of Phonetic Sciences*, Oxford: Blackwell. Ed: (Hardcastle and Laver) .PP. 371 –404.

40- Fashal, M. (1991) *Acoustic Analysis of Tempo in News-Bulletins*. PH.D. thesis. Alexandria University.

41- Fashal, M. (2006) Predicting prosodic boundaries in Arabic. Department of phonetics and linguistics, Faculty of Arts, Alexandria University. Bulletin of the Faculty of Arts, issue No. 19, vol. (56) 2006.

42- Fashal, M. and Mansour, M. (2007) The perceptual and linguistic model of pause boundaries in Classical Arabic. Department of phonetics and linguistics, Faculty of Arts, Alexandria University. Bulletin of the Faculty of Arts, issue No. 21, vol. (57) 2007.

43- Fernandez S. S. Eugenia. (2012) Glottal source parameters for forensic voice comparison: an approach to voice quality in twins voices. Phonetics Laboratory, Consejo Superior de Investigaciones Científicas, Madrid, Spain *IAFPA*,eugenia.sansegundo@cchs.csic.es.

44- Fletcher, J. (1985) Some effects of tempo change on timing in French. *Linguistics*, pp. 45-64.

45- Fors, K. L. (2011) An investigation of intra-turn pauses in spontaneous speech. 1104-5787, Institutionen för filosofi, lingvistik och vetenskapsteori, Proceedings from Fonetik 2011: TMH-QPSR, 51 s. 65-68.

46- Fougeron, C. and Ah Jun, S. A. (1998) Rate Effects on French Intonation: Prosodic Organization and Phonetic Realization. *Journal of Phonetics* 26, 45-69.

47- Foulkes, P. y. and Barron, A. (2000) Telephone speaker recognition amongst members of close social network. *International Journal of Speech, Language and the Law*, 7, 180-198.

48- Furui, S. (2008) Speaker Recognition. Tokyo Institute of Technology. Scholarpedia, 3(4):3715. doi:10.4249/scholarpedia.3715. Website: http://www.scholarpedia.org/article/Speaker_recognition

49- Gay, T. (1978) Effect Of Speaking Rate On Vowel Formant Movements. *J Acoust. Soc. Am*. 63, pp. 223-230.

50- Ghazali, S., Hamdi, R. and Barkat, M. (2002) Speech Rhythm Variation in Arabic Dialects. Proc. of Speech Prosody Aix-en-Provence, France, 331-334.

51- Gold, E. (2012) Articulation Rate as a Discriminant In Forensic Speaker Comparisons. *UNSW Forensic Speech Science Conference* Sydney, Australia. Website: http://sydney2012.forensic-voice-comparison.net/

https://phonetics-acoustics.blogspot.com

52- Grosjean, F. and Lane, H. (1974) Effects of two temporal variables on the listener's perception of reading rate. *Journal of Experimental Psychology, 102*(5), 893-896.

53- Halle, P. (1998) A Preliminary Study of Speech Rate Perception. Labo de psychlogie experimente. CNRS, EHSS, Paris, France. Se. 61.1.1- 61.1.4.

54- Hardcastle, W. J. and Laver, J. (1997) *The Handbook of Phonetic Sciences.* Oxford, Blackwell.

55- Haselager, G.J.T., Slis, I.H. and Rietveld, A.C.M. (1991) An alternative method of studying the development of speech rate. Clinical Linguistics & Phonetics 5 (1), pp. 53-63.

56- Hayward, K. (2000) Experimental Phonetics: An Introduction. Harlow: Longman.

57- Heffernan, K. M. (2007) *Phonetic Distinctiveness as a Sociolinguistic Variable.* Ph.D. dissertation,University of Toronto.

58- Hei-Yin, W. (2004) *Forensic Voice Identification under Optimum Environment, Noisy Environment and Over Telephone Condition*. PHD dissertation. *Speech and Hearing Sciences*, the University of Hong Kong, May 7, 2004.

59- Hicks J.W. (1979) *An Acoustical/Temporal Analysis Of Emotional Stress In Speech* .PhD dissertation, University of Florida

60- Hirata, Y. and Tsukada, K. (2004) Effects of speaking rate and vowel length on formant movements in Japanese. Phonetica, 66 (3):129–149.*Proceedings of the 2003 Texas Linguistics Society Conference*, ed. Augustine Agwuele, Willis Warren, and Sang-Hoon Park, 73-85. Somerville, MA: Cascadilla Proceedings Project.

61- Hollien, H. and Reva, S. (2001) Speaker identification utilizing non contemporary speech. *Journal of Forensic Science*, 46(1): 63-67.

62- Hollien, H., Liljegren, K., Martin, CA. and De Jong G. (2001) Production of Intoxication States by Actors-Acoustic and Temporal Characteristic. Copyright © 2001 by ASTM International.

63- Huggins, A. W. F. (1972) On the Perception of Temporal Phenomena in Speech. *J. Acoust. Soc. Amer.* Vol. 51. Pp. 1279-1290.

64- Igarashi, Y. (2004) "Segmental anchoring" of F0 under changes in speech: Evidence from Russian. *Proceedings of the International Conference: Speech Prosody 2004* (SP2004): 25-28. (March 23, 2004, Nara-Ken New Public Hall, Nara, Japan).

65- Jacewicz, E., Fox, R. A., O'Neill, C. and Salmons, J. (2009) Articulation rate across dialect, age, and gender. Language Variation and Change, 21 (2009), 233–256. Printed in the U.S.A. © 2009 Cambridge University Press 0954-3945/09 $16.00 doi: 10.1017/S0954394509990093.

66- Jessen M. and Bundeskriminalamt BKA (2008) Forensic Phonetics. *Language and Linguistics* Compass 2/4 (2008): 671–711.

https://phonetics-acoustics.blogspot.com

67- Jessen, M. (2007) Forensic Reference Data on Articulation Rate in German. *Science and Justice*, 47.50–67.

68- Johnson, K., & Azara, M. (2000) The perception of personal identity in speech: Evidence from the perception of twins' speech. Department of Linguistics, Ohio State University, 222 Oxley Hall. 1712 Neil Ave. Columbus, OH 43210-1298.

69- Kassem, E. M. (2014) *Contributions of nasals and vowels to speaker identification in Arabic*. PHD thesis, Alexandria University, Faculty of Arts, Department of phonetics and linguistics.

70- Kinoshita, Y. (1998) Japanese forensic phonetics: non-contemporaneous *within*-speaker variation in natural and read-out speech. *The 5th International Conference on Spoken Language Processing*, Incorporating *The 7th Australian International Speech Science and Technology Conference*, Sydney Convention Centre, Sydney, Australia, 30th November - 4th December. http://www.isca-speech.org/archive/icslp_1998/i98_0652.html.

71- Koenig, B. J. (1986) Spectrographic Voice Identification: A Forensic Survey. Letter to the editor of JASA 79/6: 2088–90.

72- Kohler, K. J. (1986) Parameters of speech rate perception in German words and sentences: duration, F_0 movement and F_0 level. Institut für Phonetik, University of Kiel. Language & Speech, 29: 115–139.

73- Koreman, J. (2006) The role of articulation rate in distinguishing fast and slow speaker. Institute of Phonetics Saarland University, Germany. jkoreman@coli.uni-saarland.de.

74- Kowal, S., O'Connell, D. C., and Sabin, E. J. (1975) Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research*, 4, 195-207.

75- Krauss, R.M., Freyberg, R., & Morsella, E. (2002) Inferring speakers' physical attributes from their voices. Department of Psychology, Columbia University, 1190 Amsterdam Avenue, New York, NY 10027, USA. *Journal of Experimental Social Psychology* 38 (2002) 618–625. www.academicpress.com.

76- Künzel, H. J. (1987) Sprechererkennung: Grundzüge forensischer Sprachverarbeitung, Heidelberg: Kriminalistik Verlag.

77- Künzel, H. J. (1997) Some general phonetic and forensic aspects of speaking tempo. Forensic Linguistics, 4.48–83.

78- Künzel, H. J. (1997) Some general phonetic and forensic aspects of speaking tempo. FL 4/1:48–83

79- Künzel, H. J. (2000) Effects of voice disguise on speaking fundamental frequency. Forensic Linguistics 7.149–79.

80- Künzel, H. J., Masthoff, H. R., & Koster, J. P. (1995) The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition.

https://phonetics-acoustics.blogspot.com

Phonetics Institute, University of Trier, Trier, Germany. *Science & Justice* 1995; 35: 291-295 (Received 23 May 1994; accepted 17 March 1995).

81- Lane, H. and Grosjean, F. (1973) Perception of reading rate by listeners and speakers. *Journal of Experimental Psychology*, *97*(2), 141-147.

82- LaRiviere, C.L. (1975) Contribution of Fundamental Frequency and Formant Frequencies to Speaker Identification, Phonetica, 31,185-197.

83- Lass, N. J. Mertz, P. J. & Kimmel, K. L. (1978) The effect of temporal speech alterations on speaker race and sex identification. *Language and Speech*, 21, 279-290.

84- Laver, J. M. D. (1994) *Principles of Phonetics.* Cambridge: Cambridge University Press.

85- Lee, S., Yildirim, S., Kazemzadeh, A., and Narayanan, S. S. (2005) An articulatory study of emotional speech production. *Proceedings ofInter Speech* (pp. 497-500). Lisbon, Portugal.

86- Lee, S., Yildirim, S., Kazemzadeh, A., and Narayanan, S.S. (2005) An articulatory study of emotional speech production. *Proceedings of Inter-Speech* (pp. 497-500). Lisbon, Portugal.

87- Leemann, A., Dellow, V. & Kolly, M. J. (2012) Exploring speech temporal features of twins: the case of V%. Phonetisches Laboratorium der Universität Zürich, presented at the *Annual Phonetics and Phonology meeting in Jena/Germany* 2012 adrian.leemann@pholab.uzh.ch , volker.dellwo@uzh.ch, marie-josé.kolly@uzh.ch.

88- Lehiste, I. (1970) Suprasegmentals, Cambridge, MA: MIT Press.

89- Li, A. and Zu, Y. (2008) Speaking rate effects on discourse prosody in Standard Chinese. Proceedings of Speech Prosody 2008, Campinas, Brazil.

90- Lindh, J. (2004) Handling the “Voiceprint” Issue. In Proceedings, FONETIK 2004, Dept. of Linguistics, Stockholm University.

91- Lindh, J. (2006) Preliminary F0 statistics and forensic phonetics. Goteborg University, Sweden. jonas.lindh@ling.gu.se.

92- Lindh, J. (2009) *Robustness of recognition of voices, speech and speakers through a forensic perspective - tools and methods.* A PhD Thesis EMBRYO, University of Gothenburg, Sweden, ISBN 978-91-977196-7-4, 2.

93- Lisker, L. (1989) On time and timing in speech. In T.A. Sebeok (Ed.), *Current Trends in Linguistics*, Vol. 12, The Hague: Mouton, 1974, PP. 2387-2418.

94- Loakes, D. (2006) *A Forensic Phonetic Investigation into the Speech Patterns of Identical and Non-Identical Twins*. PhD Dissertation, University of Melbourne.

95- Magen, H. S. and Blumstein, S. E. (1993) Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics*, 21, 387-409.

https://phonetics-acoustics.blogspot.com

96- Malécot, A., Johnston, R. and Kizziar, P.A. (1972) Syllabic rate and utterance length in French. Phonetica 26, pp. 235-251.

97- Masaki, S., Shirai, K. and Kiritani, S. (1986) Effects of tempo and context on jaw openings for vowel in vowel sequence words. *Acoustics, Speech, and Signal Processing, IEEE International*, Tokyo. P.P: 1629 - 1632.

98- McClean, M. D., & Clay, J. L. (1995) Activation of lip motor units with variations in speech rate and phonetic structure. *Journal of Speech & Hearing Research*; vol. 38(4) Aug 1995, 772-782.

99- McDougall K. & Nolan F. (2007) Discrimination of speakers using the formant dynamics of /uː/ in British English. Department of Linguistics, University of Cambridge, kem37@cam.ac.uk, Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6–10 August 2007, 1825–1828, http://www.icphs2007.de/conference/Papers/1567/1567.pdf.

100- McDougall, K. (2004) Speaker-specific formant dynamics: an experiment on Australian English /ai/. *International Journal of Speech, Language and the Law*, 11.103–30.

101- McDougall, K. (2006) Dynamic features of speech and the characterization of speakers: towards a new approach using formant frequencies". *International Journal of Speech, Language and the Law*, 13.89–126.

102- Mcpeek, T. (2013) *American Voice Types: Towards A Vocal Typology for American English*. PHD dissertation. University of Florida 2013.

103- Mikos, V.A. and Gelfer, M. P. (2001) The Relative Contribution of Speaking Fundamental Frequency and Formant Frequencies to Gender Identification of Biological Males, Biological Females, and Male-to-Female Transgendered Individuals Based on Isolated Vowels. A Paper Presented at the 2001 Convention of the American Speech-Language-Hearing Association November 15-18, 2001 New Orelans, LA.

104- Miller, J. L., and Baer, T. (1983) Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America*, 73, 1751-1755.

105- Miller, M. (1984) On the perception of rhythm. *Journal of Phonetics*, 12, 75-83.

106- Miller, M., Grosjean, F. & Lomanto, C. (1984) Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. Phonetica 41, pp. 215- 225.

107- Morris, P. S. F. (2011) Contribution of voice fundamental frequency and formants to the identification of speaker's gender. A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2011. URL: http://hdl.handle.net/10722/192901

108- Nolan F., McDougall K., Jong D. G. & Hudson T. (2006) A Forensic Phonetic Study of 'Dynamic' Sources of Variability in Speech: The DyViS Project. Department of Linguistics, University of Cambridge, United Kingdom, fjn1@cam.ac.uk , kem37@cam.ac.uk , gd288@cam.ac.uk , toh22@cam.ac.uk.

https://phonetics-acoustics.blogspot.com

109- Nolan, F. & T. Oh. (1996) Identical twins, different voices. Forensic Linguistics 3, 39-49.

110- Nolan, F. (1997) Speaker recognition and forensic phonetics. In W. J. Hardcastle & J. Laver (Eds.), the *handbook of phonetic sciences*, 744–767. Oxford: Blackwell.

111- Nolan, F. (2001) Speaker identification evidence: its forms, limitations and roles, in Proceedings of *the Conference Law and Language*: Prospect and Retrospect. 12-15 Levi, Finnish Lapland. nolan2001proclanglaw.pdf

112- Nolan, F. and Catalin G. (2005) A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law*, 12 (2): 143-173.

113- Nooteboom, S. (1997) The prosody of speech: melody and rhythm. *The Handbook of Phonetic Sciences*, Oxford: Blackwell. Ed: (Hardcastle and Laver) .PP. 256 –297.

114- O'Neill, C.G. (2008) Dialect Variation in Speaking Rate. A Senior Honors Thesis. Presented in Partial Fulfillment of the requirements for graduation with research distinction in Speech and Hearing Sciences in the undergraduate colleges of .The Ohio State University. June 2008. Project Advisors: Dr. Robert Fox and Dr. Ewa Jacewicz.

115- Ostry, D. J. and Munhall, K.G. (1985) Control of rate and duration of speech movements. *Journal of the Acoustical Society of America*, Vol. 77, pp. 640-648.

116- Pasich, C. (2006) Key Problems in Speaker Identification. Version 1.2: Dec 20, 2006 11:53 am US/Central.

117- Perrachione, T. K, Del Tufo, S. N. and Gabrieli, J. D. E. (2011) Human voice recognition depends on language ability. Science 333(6042):595.

118- Pfitzinger, H.R., (1996) Two Approaches to Speech Rate Estimation. Proc. SST, pp. 421–426, Adelaide, Australia, pp. 421-426.

119- Pisoni, D.P. (1993) Long-term memory in speech perception: some new findings on talker variability, speaking rate and perceptual learning. Speech Communication 13, pp. 109-125.

120- Recasens, D. (1985) Coarticulatory patterns and degrees of co-articulatory resistance in Catalan CV sequences. *Journal of Language and Speech*, 28, 97–114.

121- Roberts, L. (2007) Authentic and Acted Responses to Physical and Emotional Distress. the Unversity of York, lsr501@york.ac.uk.

122- Rodman, R., McAllister, D., Bitzer, D., Cepeda, L., & Abbitt, P. (2002) Forensic speaker identification based on spectral moments. *International Journal of Speech Language and the Law*, 9(1), 22-43. Website: http://research.csc.ncsu.edu/speakerrecognition/Papers/SpeakerID.pdf

123- Romito, L., Lio, R. & Galata, V. (2005) Fluency articulation and speech rate as new parameters in the speaker recognition. Article presented at *the 3rd conference on experimental phonetics (CEP)*, Santiago de Compostela.

https://phonetics-acoustics.blogspot.com

124- Rose, P. (2002) *Forensic Speaker Identification.* London, UK: Taylor and Francis.

125- Rose, P. (2006) The intrinsic forensic discriminatory power of diphthongs. Proceedings of the 11th Australasian. *International Conference on Speech Science and Technology*. 6-8 December 2006, University of Auckland, New Zealand. 64 69.

126- Rose, P. (2006) The intrinsic forensic discriminatory power of diphthongs. En P. Warren y C. I. Watson (Eds.), Proceedings of *the 11th Australian International Conference on Speech Sciences and Technology* pp.492-497). Auckland: University of Auckland.

127- Rose, P., Kinoshita, Y. and Alderman, T. (2006) Realistic extrinsic forensic speaker discrimination with the diphthong /aI/. Proceedings of *the 11thAustralasian International Conference on Speech Sciences and Technology* (pp.329-334). Sydney: University of Sydney.

128- Rose, P., Takashi O. and Kinoshita, Y. (2003) Strength of forensic speaker identification evidence: multi speaker formant and cepstrum-based segmental discrimination with a Bayesian likelihood ratio as threshold. *International Journal of Speech, Language and the Law* 10.179–202.

129- Rosen, S. (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos, Trans. R. Soc. Lond. B Biol. Sci. 336, 367–373. doi: 10.1098/rstb.1992.0070.

130- Saeed, K. (2008) Sound and Voice Verification and Identification: A Brief Review of Toeplitz Approach. Computer Engineering Department, Faculty of Computer Science.

131- Saeed, K. A. (2006) Note on Biometrics and Voice Print: Voice-Signal Feature Selection and Extraction – A Burg-Toeplitz Approach. Faculty of Computer Science, Bialystok Technical University, Wiejska 45A, 15-351 Bialystok, Poland. aida@ii.pb.bialystok.pl. http://aragorn.pb.bialystok.pl/~zspinfo/

132- Salem, H. (2006) Effects of speaking rate on the production of singleton and geminate consonants in Colloquial Egyptian Arabic. Department of phonetics and linguistics, Faculty of Arts, Alexandria University. Bulletin of the Faculty of Arts, issue No. 5, vol. (56) 2006.

133- Salleh, A. (2008) No Such Thing as a Voice Print. ABC Science Online.

134- Schelten, C. S. (2007) The Significance of Speaking Rate in Speech Treatment. Die Sprachheilarbeit, 4, 136 - 145.

135- Schiller, N. O. and Koster, O. (1996) Evaluation of a foreign speaker in forensic phonetics: a report. Forensic Linguistics, 3(1), 176-85.

136- Schwab, S. & Avanzi, M. (2015) Regional variation and articulation rate in French. *Journal of Phonetics*, 48, 96-105.

137- Singh, N., Khan, R. A. and Shree, R. (2012) Applications of Speaker Recognition. *International conference of modeling, optimization and computing (ICMOC 2012)*, Procedia Engineering Conference 2012.

https://phonetics-acoustics.blogspot.com

138- Slis, I.H., Haselager, G.J.T. & Rietveld, A.C.M. (1991) Development of Speech Rate in Dutch Children. *Clinical Linguistics & Phonetics.*

139- Smiljanic, R, and Bradlow, A. R (2008) Stability of temporal contrasts across speaking styles in English and Croatian, *Journal of Phonetics,* 36, 91-113.

140- Starkweather, C. W. (1987) Fluency and stuttering.Englewood Cliffs, *NJ*: Prentice Hall.

141- Stevens, K. N. (1998) *Acoustic phonetics.* Cambridge, MA: MIT Press.

142- Strik, H., Doremalen,J. V. and Cucchiarini, C. (2008) Pronunciation reduction: how it relates to speech style, gender, and age. *Proceedings of Interspeech-2008*, *Brisbane, Australia*, Sept. 26-29, 2008, pp. 1477-1480.

143- Tillmann, G. H. and Pfitzinger, R. H. (2003) Local speech rate: relationships between articulation and speech acoustics. Department of phonetics and speech communication, German University.

144- Trouvain, J. (2003) *Tempo variation in speech production, implications for speech synthesis*. PH.D dissertation.

145- Trouvain, J. and Grice, M. (1999) The effect of tempo on prosodic structure. Proc. ICPhS (2), San Franscisco, pp. 1067-1070.

146- Vaane, E. (1982) Subjective Estimation of Speech Rate. Phonetica 39, pp. 136- 149.

147- Van Son, R.J.J.H. and Pols, L.C.W. (1992) Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of Acoust.Soc.Am.* 92, 121-127.

148- Vingerling, M. (1984) Speech rate and its differential influence on the perception of normally hearing and hearing impaired subjects. Institute of Phonetics, University of Utrecht. PP. 778 -781.

149- Watt, D. (2010) The identification of the individual through speech. Eds: (En C. Llamasy D. Watt) chapter 8, *Language and Identities* (pp. 76-85). Edimburgo: Edinburgh University Press.

150- Whiteside, S.P. and Hodgson, C. (2000) Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. Speech Communication 32 (4), pp. 267-285.

151- Wilson, C.Y. E., Manocha, S. and Vishnubhotla, S. (2006) A new set of features for text-independent speaker identification. In Proceedings of ICSLP.

152- Yarmey A. D. (2001) Ear witness descriptions and speaker identification. *The International Journal of Speech, Language and the Law*, Vol 8 No 1 (2001). http://www.equinoxjournals.com/ojs/index.php/IJSLL.

153- Zellner, B. (1994). Pauses and the temporal structure of speech, in E. Keller (Ed.) *Fundamentals of speech synthesis and speech recognition*. (pp. 41–62). Chichester: John Wiley.

https://phonetics-acoustics.blogspot.com

154- Zhang, Y., Nissen, S. L., and Francis, A. (2008) Acoustic characteristics of English lexical stress produced by Mandarin speakers. *Journal of the Acoustical Society of America*, 123(6), 4498- 4513.

Online Journals:

- International Journal of Speech, Language and the Law: http://www.equinoxjournals.com/ojs/index.php/IJSLL.
- Journal of Voice: http://www.jvoice.org/
- Journal of Acoustical Society of America: http://asadl.org/jasa/
- Journal of Phonetics: http://www.journals.elsevier.com/journal-of-phonetics/

Online websites:

- http://psycnet.apa.org/psycinfo/1994-12862-001
- http://www.digisoundmastering.com/forensic-audio/forensic-speaker-identification
- http://www.doc.ic.ac.uk/~nd/surprise_95/journal/vol4/sm1/report.html
- http://www.docstoc.com
- http://www.documbase.com
- http://www.ling.cam.ac.uk/gea/forensicphonetics/forensicphonetics.htm
- http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2790192/
- http://www.owlinvestigations.com/article1.html
- http://www.praat.org
- http://www.scholarpedia.org/article/Speaker_recognition
- http://www.sciencedirect.com
- http://www.sciencemag.org
- http://www.springer.com
- http://www.ufdc.ufl.edu/
- http://www.ukessays.com/essays/sciences/forensic-speaker-identification.php
- https://www.boundless.com/
- The International Association of Forensic Phonetics and Acoustics (2004): http://www.iafpa.net

https://phonetics-acoustics.blogspot.com

مكتبة وملتقى علم الأصوات
phonetics-acoustics.blogspot.com

التعرف على المتكلم اعتماداً على معايير السرعة الزمنية

هدف هذه الدراسة هو التعرف على هوية المتكلمين غير المعروفين من سرعة كلامهم، وقد تم فى هذا البحث - **على المستوى الإدراكي -** عمل تقييم للقدرة الإدراكية للمستمعين غير المدربين فى التعرف المتكلم اعتمادا على سرعة كلامه، وإدراك ما إذا كان الأسرع أم الأبطأ بين المتكلمين العشرة الذين تم اختيارهم للتجربة. أما **على المستوى الأكوستيكي،** فقد تم رصد المعايير الفيزيائية الأساسية للتعرف على صوت المتكلم وهى كالآتى:

1. التردد الأساسى F_0
2. الترددات المكونة للصوائت(F1, F2, F3).
3. الرنين الأنفي للصوامت (الغنة).
4. معدل سرعة الكلام (SR) ومعدل سرعةالمنطوقات (AR)

وقد اختير العنصر الأخير وهو سرعة الكلام (Speech Tempo) كموضوع لهذه الدراسة. وقد تم عمل التحليل الفيزيائى لكلام المتحدثين وقياس معدل سرعة الكلام ومعدل سرعة المنطوقات والوقفات فى كلام كل متحدث (أطوالهم وأعدادهم). هذا فضلا على قياس التردد الأساسى لكل متكلم F_0 وشدة الصوت (I).

هناك العديد من الأسباب الأساسية التي توضح مدى أهمية المعايير الزمنية ومعدل سرعة الكلام فى التعرف على المتكلم للأغراض القضائية وهى كالآتى:

1. لا يمكن محاكاة سمات السرعة الزمنية للكلام.
2. لا يمكن للمتكلم السيطرة على السرعة الزمنية لكلامه بشكل واع.
3. الفروق الفردية بين المتكلمين تُعد من أهم مصادر التغير التي تُؤثر على معدل سرعة الكلام.

<u>المنهجية البحثية</u>

تشمل هذة التجربة عشرة أشخاص (خمس نساء وخمسة رجال) غير معروفين الهوية ومتحدثين أصليين للهجة العامية العربية وتقدر أعمارهم بين 19 و 40 عام. تتكون المادة من كلام تلقائي لمدة نصف دقيقة (30 ثانية) لكل متكلم مع تجنب تأثير أو سيطرة أي نوع من أنواع المشاعر السلبية للمتكلمين.

تم تسجيل المادة من خلال برنامج "الصحافة فى عيونهم" الذي يذاع يومياً على راديو إذاعة الإسكندرية. وتم تحليل المادة المسجلة لكل متكلم يدويا و كتابتها بالرموز الصوتية Transcription وذلك عن طريق الإستماع الجيد لهذه المادة المسجلة مراراً وتكراراً بواسطة Praat Software. ثم تمت عملية فصل المقاطع Segmentation Process وذلك لحساب SR & AR. كما تم أيضا قياس التردد الأساسي و شدة الصوت لكل متكلم ودرجة التلعثم و عدد الوقفات وزمن كل وقفة ونوعها ونسبة كل الوقفات إلى مدة الكلام الكاملة .

<u>الإختبار الإستماعي للمستمعين</u>

ستون مستمع من طلبة الجامعات ومتحدثين أصليين أيضاً للعامية العربية المصرية وتتراوح أعمارهم بين 17 و25 عام ، جميعهم تطوعوا للإشتراك فى هذا الإختبار. المهمة الأساسية للمستمعين هى الإستماع بحرص شديد إلى المتكلمين العشرة وتحديد المتكلم الأسرع و أيضا المتكلم الأبطأ من حيث سرعة الكلام عن طريق وضع علامة (√) أمام الرمز الدال عليه.

<u>النتائج</u>

تشير النتائج إلى أن:

1. أكوستيكياً و إدراكياً: سرعة الكلام موضحة فى معدل سرعة الكلام (SR)هى المعيار الأقوى فى التعرف على المتكلمين غير المعروفين; بينما معدل سرعة نطق الأصوات (الصوامت والصوائت) (AR) كان أقل تأثيرا على تحديد سرعة المتكلم.
2. النسبة المئوية للوقفات تلعب دوراً مهما جداً على المستويين الإدراكي والأكوستيكي؛ على المستوى الإدراكي فإن زيادة النسبة المئوية للوقفات تُعدّ من أهم العناصر التي تؤثر على إدراك المستمعين للسرعة الزمنية للكلام، حيث تشير إلى سرعة الكلام البطيئة. أما على المستوى الأكوستيكي: فليس لها أي تأثير واضع على زيادة أو نقصان سرعة الكلام للمتكلم.
3. درجة التلعثم فى الكلام (الوقفات المملوءة filled pauses)، تُعد من العناصر المُميزة فى التعرف على المتكلم الأسرع من حيث سرعة الكلام للمتكلم. ومع ذلك فليس لها أي دور فعّال فى التعرف على المتكلم الأبطأ .
4. التردد الأساسي للمتكلم يُعد من العناصر الأكوستيكية المُميزة لتحديد سرعة الكلام للمتكلم ،بحيث زيادة التردد الأساسي للمتكلم تشير إلى زيادة معدل سرعة كلامه إدراكيا وأكوستيكيا.
5. متوسط شدة الصوت لدى المتكلم يُعد من الناحية الإدراكية من العناصر المُميزة بالنسبة إلى آذان المستمعين، بحيث زيادة شدة الصوت تشير إلى زيادة معدل سرعة الكلام للمتكلم ,وأيضا نقصان شدة الصوت تدل على نقصان معدل سرعة الكلام للمتكلم. ولكن هذه النتائج لا تنطبق على المستوى الأكوستيكي.

https://phonetics-acoustics.blogspot.com